DOCUMENT RESUME

ED 265 070 SE 046 371

AUTHOR Zehavi, Nurit
TITLE Investigation of the Effect of Assignment Projects on Mathematical Activity of Graduating Junior High School Students.
PUB DATE Nov 84
NOTE 238p.; Ph.D. Thesis, The Weizmann Institute Science, Rehovot, Israel.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS Doctoral Dissertations; Foreign Countries; *Grade 9; Junior High Schools; Junior High School Students; *Learning Activities; *Mathematics Instruction; *Problem Solving; *Secondary School Mathematics
IDENTIFIERS Israel; Mathematics Education Research

ABSTRACT

This study explored student mathematical activity in open problem-solving situations, derived from the work of Polya on problem solving and Skemp on intelligent learning and teaching. Assignment projects with problems for ninth-grade students were developed, whether they elicit the desired cognitive and cogno-affective goals was investigated, and their effectiveness in enhancing mathematical activity was evaluated. Student responses were analyzed on quality-of-answer and quality-of-procedure scales. Improvement is exhibited in a quality of solution table. Assignment tests were developed, validated, and used to compare changes in mathematical activity in a pre-post control-experimental setting. The results indicate that the assignment projects were effective. It also became clear that teachers needed more support in a practical form. A clinical study was undertaken; students in one class were interviewed about their work on assignments. The researcher's premise was that if he could categorize student difficulties and the guidance into some sort of hierarchy, then he would have a model and a language which would help teachers in their work. The applicability of the diagnostic guidance hierarchy was examined successfully in a regular classroom setting. (MNS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED265070

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

INVESTIGATION OF THE EFFECT OF [illegible]
ON MATHEMATICAL ACTIVITY OF [illegible] HIGH SCHOOL STUDENTS

Thesis
for the degree of
DOCTOR OF PHILOSOPHY

by

Nurit Zehavi

Submitted to the
Scientific Council of the
The Weizmann Institute of Science
Rehovot, Israel

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nurit Zehavi
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 046 371

INVESTIGATION OF THE EFFECT OF ASSIGNMENT PROJECTS ON MATHEMATICAL ACTIVITY OF GRADUATING JUNIOR HIGH SCHOOL STUDENTS

ABSTRACT

The object of the work described in this thesis is to explore and enhance student mathematical activity in open problem solving situations. The initial efforts drew their inspiration from Polya's work on heuristic learning. The learning models have their source in Skemp's ideas of intelligent learning and teaching.

In the first place, *assignment projects* for 9th grade students were developed within a framework of cognitive and cogno-affective goals. An assignment project is an open-ended mathematical problem, whose scope is wide but is integral to the core curriculum.

Studies were carried out to investigate whether in fact the material elicits the desired goals. This was followed by an evaluation of its effectiveness in enhancing student mathematical activity.

Based on the Skemp model a framework for qualitative evaluation of mathematical activity was developed. Student responses were analysed on a *quality-of-answer* scale and a *quality-of-procedure* scale. Quality of procedure, in principle, may limit the quality of answer. However a student may arrive at a general pattern of answer without a proof, or a high order procedure may not lead to the full answer. The two dimensions - quality of answer and of procedure - can be exhibited in a *quality-of-solution* table. Improvement of mathematical activity is exhibited in such a table by a "population shift" in the desired direction. The statistical significance of the improvement can be tested, using the Wilcoxon-Raatz test, which is a modified version of the nonparametric

Wilcoxon matched-pairs signed-rank test. *Assignment tests,* within this framework were developed and validated in a series of studies. They were used to compare changes in mathematical activity in a pre-post-control-experimental setting. The results indicate that the *assignment projects* are indeed effective.

During these studies it became clear that teachers needed more support in a practical form. Therefore, a clinical study was undertaken, in which one class was chosen and all the students interviewed about their work on assignments within the goal-framework. The premise was, that if we could categorize student difficulties and the guidance into some sort of hierarchy, and establish a link between them, then we would have a model and a language which would help teachers in their work. This led to the establishment of a hierarchy of guidance. Finally, the applicability of the diagnostic guidance was examined successfully in regular classroom setting.

To my family

This study was carried out in
The Department of Science Teaching
The Weizmann Institute of Science

under the supervision of

Prof. M. Bruckheimer

and

Dr. Ruth Ben-Zvi

6

ACKNOWLEDGEMENTS

My sincere appreciation is extended to Professor Maxim Bruckheimer for his most inspiring supervision of this research at every stage.

I am very grateful to my other supervisor Dr. Ruth Ben-Zvi for her counsel and most helpful constructive criticism.

I wish to thank my colleagues at the Mathematics Group for their stimulating cooperation in the development of the assignment projects and for their support in the evaluation process. Special thanks to Mrs. Yetty Varon who provided expert assistance in statistics.

I would like to express my warm gratitude to Dr. Bat-Sheva Eylon for her invaluable help in a critical stage of this work.

I would further like to thank all the teachers and students who participated in the study for their enthusiastic involvement.

Thanks to the staff of the Department of Science Teaching, especially to Mrs. Yvette Savariego for patiently typing this thesis.

> *The really interesting problems of education are hard to study. They are too long-term and too complex for the laboratory, and too diverse and non-linear for the comparative method. They require longitudinal study of individuals, with intervention a dependent variable, dependent upon close diagnostic observation. The investigator who can do that and will do it is, after all, rather like what I have called a teacher. So the teacher himself is potentially the best researcher, if only we would offer him strong intellectual support and respect his potentialities as a scientist: lighten his mechanical burdens, join him more frequently in his association with children, argue with him, pick his brains.*
>
> David Hawkins (1973)

CONTENTS

1. Introduction 1

  1.1. Rationale 1

  1.2. High level activity: development, evaluation, implementation 1

  1.3. Theoretical boundaries 4

  1.4. Review of relevant research instruments 7

2. The study 11

  2.1. Methodology 11

  2.2. Outline of the research 12

  2.3. Objectives of the field research studies 13

3. Development related to goals 14

  3.1. Goals of the *assignment projects* 14

  3.2. Development of *assignment projects* 15

  3.3. An example: the project *Squares* 17

4. Investigation related to goals 19

  4.1. Descriptive analysis 19

  4.2. Qualitative analysis 24

5. Evaluation instruments 27

  5.1. General information 27

  5.2. Achievement test 28

  5.3. Attitude questionnaire 29

  5.4. Preliminary comparative study 30

5.5. Assignment test 32
    5.5.1. Design and construction 32
    5.5.2. Validation studies 37
    5.5.3. Statistical test 40
6. Comparative evaluation study 44
    6.1. The study 44
    6.2. General findings 46
    6.3. Findings on the assignment test 51
7. Diagnostic evaluation of mathematical activity 61
    7.1. Post-evaluation studies 61
    7.2. Clinical study 62
        7.2.1. Qualities of mathematical activity and guidance hierarchy 62
        7.2.2. Diagnostic evaluation 75
    7.3. Communication study 79
    7.4. Interpreting model leading to further studies 85
8. Concluding remarks 93
Bibliography 97

Appendices

Appendix A : Text of a sample of Assignment Projects 103
Appendix B : Evaluation instruments 149
Appendix C : Comparative study: Group comparison - of average scores and distribution 163
Appendix D : Comparative study: Population shift for assignments 1 and 3 175
Appendix E : Communication study: Population shift for assignment 1-3 183
Appendix F : A model for comparing teacher-expectation and student performance 195

1. INTRODUCTION

1.1. Rationale

Mathematical activity is not only content and skill based. It is made up of constituents such as investigation, problem solving, proof, generalization and extension. The whole is greater than its parts and by the composition of these aspects we mean "real mathematical activity".

In the field of mathematics teaching there is an on-going trend to progress from an intuitive art to an experimental science.

A major problem in the study of student mathematical activity in open problem solving situations, is the provision of a suitable theoretical framework against which this activity can be objectively evaluated. One approach to this problem is the "atomization" of student activity into small mathematical steps. By contrast, an alternative approach can be described as "molecularization", that is, the breaking down of the activity no further than that which may be reasonably assumed to be performed by the student in one strategic step. In the investigation of student intellectual activity in complex situations, the atomic level is less relevant than the molecular one.

In recent years researchers have attempted to develop models and measures of student activity at the molecular level. This work may be regarded as another contribution in this direction.

1.2. High level activity: development, evaluation, implementation

Bloom et al. (1956) introduced a cognitive taxonomy which became very popular in education and may have helped to attract attention to the need for activities at the higher levels of the hierarchy.

Over the past 30-40 years, we find various books and papers devoted to aspects of mathematical problem solving, which is one of the major activities at the top end of the hierarchy. Thus Hadamard (1954), Polya (1945, 1959, 1962) and Wertheimer (1959), among others, have explored the psychological aspects of high level processes. They suggest ways of encouraging them, usually by the posing of appropriate problems.

Some desirable features of "research problems" as given by Polya (1962) are:

"(1) ... the mathematician's most crucial step may be to *choose* his problem; ... the student should *have some share in proposing the problem.*

(2) ... the problems have a *rich background;* they suggest challenging problems which suggest further challenging problems ...

(3) ... observation, conjectures, inductive arguments, in short, *plausible reasoning* play a prominent role."

Polya concludes that, of the three points, the latter two are by far the most important.

Kilpatrick (1978) maintains that instructional programs that demonstrate how heuristic techniques can be used to solve complex mathematical problems and then provide guidance in applying these techniques, show promise both for research and for practice.

Halmos (1980) believes that problems are the heart of mathematics, and hopes "that as teachers, in the classroom, in seminars, and in the books and articles we write, we will emphasize them more and more, and that we will train our students to be better problem-posers and problem-solvers than we are".

The first recommendation of the NCTM (1980) is that problem solving should be the focus of school mathematics. It is also recommended that "the evaluation of mathematics programs should be based on the program's goals, using evaluation strategies consistent with the goals".

Wheeler (1978) maintains that the ability to mathematize is necessary in situations which contain challenges and unfamiliarity. By "mathematization" he means the "mental processes that bring mathematics into being." Mathematization does not take place in vacuum; there has to be some ordered knowledge and skills with which to create. Thus, since the majority of problems to be found in books such as those cited above, are isolated from the regular school curriculum, "perhaps the greatest challenge is to produce material which fosters these developments and at the same time develops the pupils' knowledge of important particular mathematical ideas" (Bell, 1979). That is, we need to create problems integrated with the regular curriculum; problems, whose essential subject matter is knowledge and skills which the student is already supposed to have, but which have characteristics cited from Polya above.

The main difficulty in the development of materials which may achieve some of these goals, is to invent activities which are rich enough and still require only techniques familiar to young students. Another problem is to determine the criteria for evaluating student work on open problems. Furthermore, there is a pedagogic challenge: how can one guide students to plan a solution to a non-routine question to which an algorithmic approach is not applicable?

Even if one *develops* material designed to enhance mathematical activity of young students, the *implementation* of such material is problematic. It requires time and work on the part of the classroom teacher, and the regular type of instruction, as well as the standard evaluation of student work and progress, cannot be employed. The difficulties that teachers have in evaluating student progress are also shared by the developers of the materials. How does one *evaluate* the efficiency ("success") of such a program? A suitable theoretical framework can be of help.

1.3. Theoretical boundaries

Beth (Beth and Piaget, 1966) notes that, despite the important role played by Platonism and the deductive method in shaping the development of mathematical thought, these are not the only factors which have played a part. Intuition, for example, has proved to be a valuable source of new ideas and has often given rise to theorems only later confirmed by formal demonstration. While Beth is interested in the historical development of logico-mathematical thought, Piaget is interested in the way in which it developes in the individual. He finds that intellectual behavior consists at first of simple classificatory and relational activities. The child compares, distinguishes and orders the objects around him, his logical and mathematical activities, in which propositional or formal operation occur, develop out of these.

Besides developmental differences, Burton (1980) lists other individual differences in the classroom: mode of thinking (inactive/ ikonic /symbolic), thinking style (visualiser/ symboliser), preferred thinking pattern (convergent/divergent), understanding (instrumental/relational), thinking strategies (memory, pattern, operational), approach (impulsive/reflective), and attitude to learning mathematics (positive/negative).

A different approach suggests that instead of looking at the solver, one look at the problem itself as a determinant of behavior. In this respect, the work on artificial intelligence and human simulation have helped to clarify some aspects of problem solving and provide models useful for further research. The work of Newel and Simon (1972) describes the major advances in this area. With the help of ideas from information processing, attempts to understand how the human mind is able to solve complex problems requiring planning and strategy have been made. Thus, Landa (1974) and Krutetskii (1976) have shown how the information processing approach can help in the identification of problem types and the processes used in solving problems of a given type.

In teaching, there is a need to combine the two aspects discussed above, the solver and the problem. There is a need for research into methods and techniques of teaching students how to solve problems.

Scandura (1977) offers a structural learning theory which is designed to be comprehensive in scope. It deals with the role of the teacher as well as the learner, and provides a unified theoretical system that incorporates content, cognition and individual differences, and their relationships.

Skemp (1979a) presents a new model of intelligence which is concerned with the ways in which intelligence acts. He introduces a two-level cybernetic model, and regards learning as a change of state of a first level director (delta-one, in Skemp's terms), toward states which make possible better functioning; the changes are goal directed and are accounted for by a second level director system (delta two). Skemp's hypothesis is that intelligent learning can be explained in terms of a second-order goal director system, delta-two, acting on a teachable first-order director system, delta-one. The schemas that are used by delta-one in its functioning are constructed in the delta-two system. Their construction is a combination of building and testing. Three modes of building and three modes of testing are distinguished.

SCHEMA CONSTRUCTION

BUILDING	*TESTING*
Mode (i) *from our own encounters with actuality: experience.*	*Mode (i)* *against expectations of events in actuality: experiment.*
Mode (ii) *from the realities of others: communication.*	*Mode (ii)* *comparison with the realities of others: discussion.*
Mode (iii) *from within, by formation of higher-order concepts, by extrapolation, imagination, intuition: creativity.*	*Mode (iii)* *comparison with one's own existing knowledge and beliefs: internal consistency.*

The three building modes can be used with any one of the three testing modes. The construction happens when "delta-two acting on delta-one takes it from a state in which delta-one has no plans for the achievement of some specified goal state, through states in which there is some kind of plan, to a goal state in which there is an optimal plan".

In the process of constructing schemas students need the guidance of a teacher who has experienced such activities. Successful cooperation between teacher and students, according to Skemp, involves a cooperation between a delta-two system of the learner and a delta-one system of the teacher. Such a cooperation is by its nature difficult, because of the different learning goals that teachers and students may have.

Skemp (1979b) distinguishes between instrumental, relational and logical understanding and the three different kinds of learning goals associated with them.

Norman (1981) describes the characteristics of a feedback model of instruction. Starting with educational goals the teacher compares those goals with the current state of classroom behavior and knowledge, and uses an instructional strategy appropriate to the situation. The teacher continually monitors classroom behavior and modifies the instructional strategy.

In any of these theoretical frameworks, the question is how actually to connect between the teacher, the student and open problems that cannot be treated algorithmically. What methods and techniques can be used to obtain what Skemp describes as successful cooperation between the teacher's delta-one system and student's delta-two system?

1.4. Review of relevant research instruments

The Bloom Taxonomy, mentioned in section 1.2, attempted to establish suitable instruments to evaluate the quality of learning. However, it came to be used mostly to prepare items, and not to evaluate open-ended responses to existing questions (Bloom et al., 1971). Further, Kilpatrick (1979) criticises the separation in the Bloom Taxonomy of the cognitive from the affective domain. He suggests the need to find new perspectives, better suited to contemporary problems of evaluation in mathematics education.

In recent years, researchers have attempted to develop models and measures of student performance which are more applicable in classroom evaluation. The Iowa Problem Solving Project (IPSP) uses a four-step model (getting to know the problem, choosing what to do, doing it and looking back) for the problem solving process (Schoen and Oehmke, 1980). The Iowa Problem Solving Test is an easily administered multiple-choice test. It provides information about problem solving skills which correlates well with data from individual interviews. The test contains three subtests related to three of the four steps in the model: getting to know the problem, doing it and looking back. The IPSP team admit that, after two years of effort, they have not been able to find a valid machine-scorable way to test student ability to choose a problem solving strategy.

Bell and Shiu (1981) report work on the development and use of test material designed to measure the progress of students following a course with particular emphasis on process aspects of mathematics. Their problem in designing process test items was that, the strategies to be evaluated, only come into play when a sufficiently complex situation has been generated. The usual short test items are thus not applicable. Hence they prepared two tests containing six items each, and two other tests containing three longer items. For example, one of the items in a six-item-test is aimed to test an aspect of *explanation;* in particular, the ability to present a complete argument. From the scripts it was not possible to assess student understanding reliably, but the degree of explicitness of the argument, and thus its quality as a communication, was assessable.

Schoenfeld (1982) is concerned with explicit heuristic training and presents measures for student activity. He uses a paper-and-pencil test which contains five problems. For each problem a list of all plausible approaches is prepared. Then for every student paper and each approach, information is gathered about "evidence" of the approach, "pursuit" of the approach and "progress" (little, some, almost, solved) toward a solution.

A recently developed taxonomy, based on the "structure of the observed learning outcome (SOLO)" (Biggs and Collis, 1982), was used as a basis for the development of a set of *superitems* (a set of test items based on a common situation) designed to assess mathematical problem-solving. Biggs and Collis began by setting themselves the task of giving teachers actual classroom examples of levels of cognitive development in student responses. They considered four stages: *sensorimotor, intuitive/pre-operational, concrete operational, formal operational*. Within each stage they classified the structure of the responses: *uni-structural* responses represent the use of only one relevant aspect; *multi-structural,* several disjoint aspects, usually in a sequence; *relational,* several aspects related into an integrated whole; *extended abstract* takes the whole process into a new mode of functioning.

Based on the SOLO taxonomy, superitems were designed to elicit increasingly complex levels of reasoning from the students (Romberg et al., 1982a, 1982b). A pool of mathematical situations and a set of items for each situation were designed to provide information about qualitatively different levels of reasoning ability.

Schoenfeld (1983) and others recommend that research instruments should be so designed that they can also be used as part of the teaching process. An example of this is the CSMS mathematics research (Hart, 1981), which involved the writing of word problems and then, the use of these problems in both paper and pencil testing and interviews with children. Hart (1983) stresses that the interviews should be structured in such a way that the teacher can replicate the type of discussion. By doing so he may possibly add to his own repertoire of teaching skills by using interview techniques.

Cobb and Steffe (1983) believe that the researcher must act as a teacher. They suggest a constructive view of teaching, in which the teachers should continually make a *conscious* attempt to "see" both their own and the students' actions from the student point of view. By teaching in this way the teacher acts as a researcher.

The above review indicates a trend to emphasize the following aspects in the research of learning processes in mathematics:

- construction of new theoretical frameworks,
- qualitative evaluation,
- complex test items,
- applicability to classroom teaching.

These aspects play a major role in the present study of high level activities in mathematics.

2. THE STUDY

2.1. Methodology

This study is an attempt to explore and enhance the mathematical activity of graduating junior high school students.

The fundamental premise is that "real mathematical activity" is composed of a variety of activities such as problem solving, proof, generalization and extension. In order to make manageable sense of such a complex situation, we defined an operational framework of learning goals. Within this framework we :

a) designed activities for students (*assignment projects*);

b) investigated whether in fact, the activities can lead to the desired goals;

c) evaluated their effectiveness in teaching (do they enhance mathematical activity of students?);

d) diagnosed student difficulties in their attempt to achieve the goals;

e) made the conclusions of the study "communicable" for practical implementation by the classroom teacher.

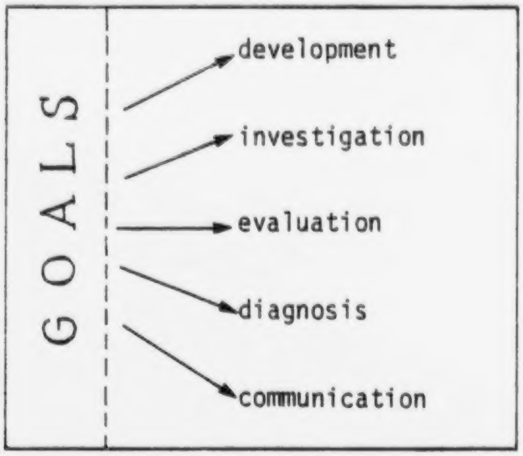
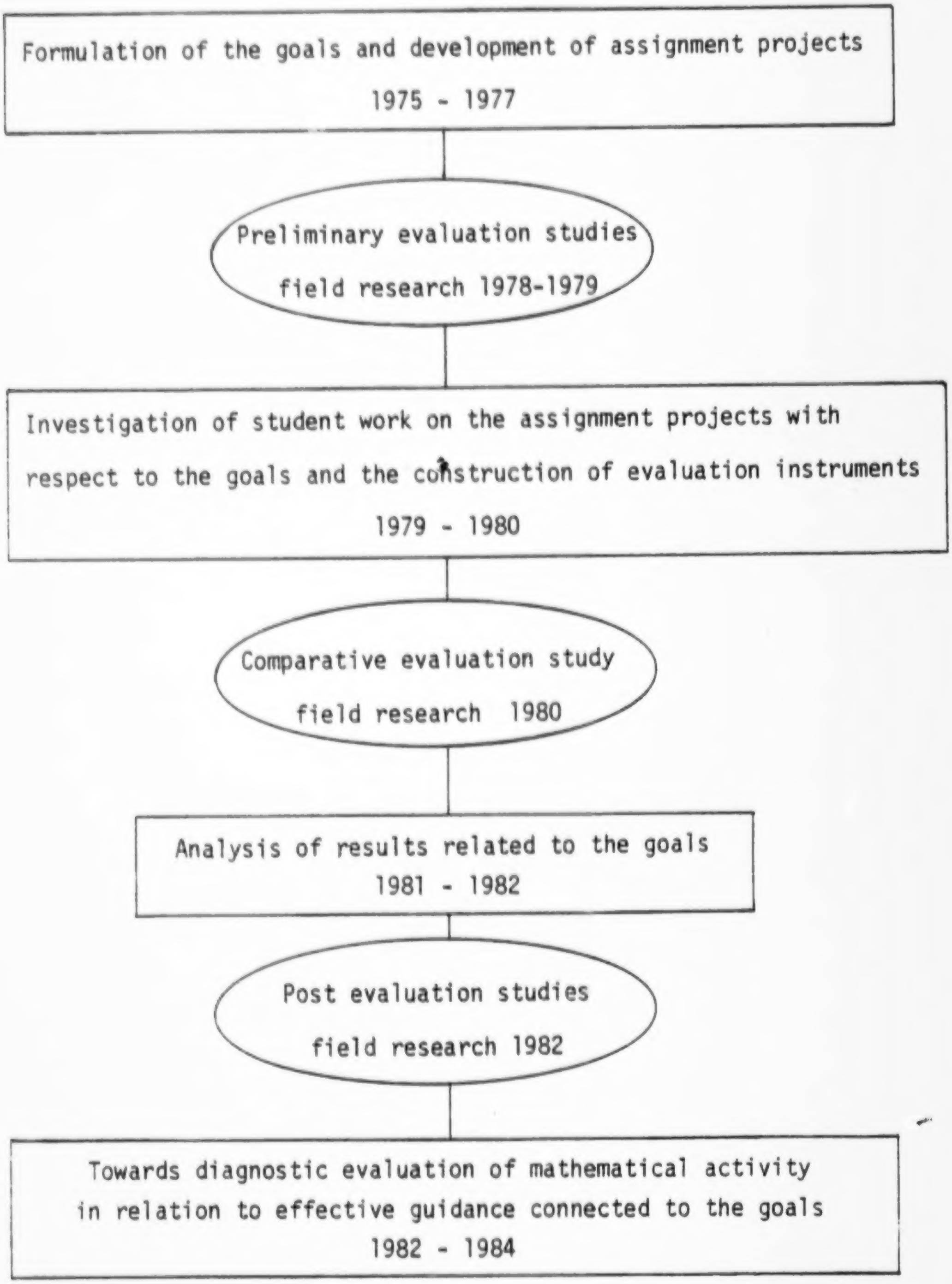


Figure 1 : Research-study within a goal-framework

2.2. Outline of the research

The various phases of the work are given in the chart below and will be discussed in the following chapters.


Formulation of the goals and development of assignment projects
1975 - 1977
Preliminary evaluation studies
field research 1978-1979
Investigation of student work on the assignment projects with respect to the goals and the construction of evaluation instruments
1979 - 1980
Comparative evaluation study
field research 1980
Analysis of results related to the goals
1981 - 1982
Post evaluation studies
field research 1982
Towards diagnostic evaluation of mathematical activity in relation to effective guidance connected to the goals
1982 - 1984

2.3. Objectives of the field research studies

The objectives of the preliminary evaluation studies were:

a) To investigate the relation between student activity and the expected goals, for those students who worked on the projects (cf. 4.1).

b) To establish a framework for qualitative analysis of student work on the projects in terms of the goals (cf. 4.2).

c) To construct and validate the research instruments (cf. 5.2, 5.3, 5.5).

d) To obtain a preliminary comparison between students who worked on the projects and students who did not (cf. 5.4).

The objective of the comparative evaluation study was:

To compare changes in mathematical activity in a pre- post-control-experiment setting (cf. 6.2, 6.3)

The objectives of the post-evaluation studies were:

a) To diagnose student mathematical activity in terms of cognitive and cogno-affective goals (cf. 7.2.1).

b) To establish a guidance system related to qualities of mathematical activity (cf. 7.2.2).

c) To test in practice the "communicability" of the diagnosis and guidance to the classroom teacher (cf. 7.3).

3. DEVELOPMENT RELATED TO GOALS

3.1. Goals of the *assignment projects*

The overall goal of the *projects* is to enhance student mathematical activity. To develop material to achieve this general goal, we need a framework of operational goals.

Under the heading of learning goals in general, and for mathematical activity in particular, we include both cognitive and affective aspects. In fact these two are not distinct and, in many cases, the affective inevitably affects the cognitive. Yet very often the affective aspect is within a subject related content. When this occurs we shall refer to *cogno-affective* goals.

Formulation of the goals

By the enhancement of student mathematical activity, we mean the improvement of student responses in the following aspects.

Non-avoidance of unfamiliar problems

Faced with a problem which appears unfamiliar, students will not avoid it, but they will make a serious attempt, either by spontaneous activation of the concepts represented by the problem or by reflective planning.

Quality of solution

In quality of solution we distinguish between:

quality of answer (A) and quality of procedure (P).

(A) - the answer will be complete; e.g. explicit consideration of all possible cases, rather than satisfaction with the first answer(s) obtained, and the establishment of a pattern if possible.

(P) - the use of higher order procedures in building and testing the answer; e.g. use of general proofs rather than particular examples and the provision of complete arguments.

Quality of procedure, in principle, may limit the quality of answer. But a student may arrive at a general pattern of answer without a proof and, on the other hand, a high order procedure may not lead to the full answer.

Extension of solution

The response to a question will go beyond the immediate and routine and will extend the answer and/or the procedure.

Extension of question

Students will pose and (try to) answer problems suggested by the given question. Their problem will be interesting and original rather than just "changing the numbers" in the given problem.

Usually students are expected to solve problems and stop. Extension of answer and/or procedure and the insertion of further questions are elements in real mathematical activity. These activities can be prompted by asking the student specifically to extend, or which is more desirable, unprompted, arising from the involvement with the given problem.

Almost any situation can be extended but, of course, there are cases where it is more likely to happen, for example, when curiosity is aroused or there is some surprise.

3.2. Development of *assignment projects*

The current educational literature is full of ideas and materials designed to encourage high level activities among students, but the suggested problem situations are often not part of a core curriculum. It seemed desirable in our view to create open-ended problem-solving situations arising from the regular junior high school mathematical lesson. We believed that such materials would be used by teachers in classes with many students, and not only in extra-curricular activities for high ability and already motivated students. Further more such situations would provide an opportunity for students to revise and extend their mathematical knowledge and skills in an interesting way. We felt that it was important (and challegding) to develop something interesting, specifically in the "dull and routine" areas of mathematics and by this enrich the program. We had all this in mind while working on the *assignment projects*.

By an *assignment project* we shall mean an open-ended mathematical problem whose scope is wide, but which is strictly related to the regular curriculum and to the core mathematics thereof. The students can, and are expected, to apply a variety of the mathematical techniques and thought processes from their experience to-date. Although based on a curriculum already familiar to the student, it has some aspect of unfamiliarity. The expectation is that the student will be encouraged by the context of the project to overcome the unfamiliarity. A typical project will be made up of a sequence of items all dealing with one situation (cf. 3.3). There will be standard and relatively technical items, intended to get the students started, but there will also be several items for which various levels of answer and procedure are possible. Some intellectual rewards are included which are intended to motivate the students.

Throughout the project, phrases like "what can you say ..." are used, indicating that more than the usual brief answer is expected. The project is opened up even further by the sections in which students are asked to invent similar problems of their own and to solve them (if possible) which is as close as we can get to real mathematical activity.

The projects were designed for ninth grade students to do on their own, each over a period of about a month, with occasional "sign posting" by the teacher. Detailed teacher-guides were prepared, in which several alternative solutions are given wherever possible, and also the mathematical and didactical background describing the purpose and construction of the projects. The development of the projects was a team effort over about three years (1975-1977). Each project was tested in teachers' workshops and with students in school and reworked by the team several times.

Examples can be found in Appendix A (text of the projects and the solution as prepared for the teachers) and in Bruckheimer and Hershkowitz (1977); Zehavi (1979); Zehavi and Bruckheimer (1982b). We discuss one project, *Squares*, the text of which can be found in Appendix A, in the following section.

3.3. An example: the project *Squares*

The project *Squares* was developed as a direct result of our wanting to create a situation in which the student "plays" with numbers, guesses and checks his guess by applying the algebraic techniques in the curriculum. Further we wanted a situation in which specification of the relevant domain (set of substitution) has meaning; hence we chose a geometric form in which to pose the problem. Also, such a setting would naturally invite further investigation of the expressions obtained.

We start from a concrete situation involving one variable, where the sides of a square are extended by 2 cm to produce a new square. Fig. 2 shows some of the possibilities

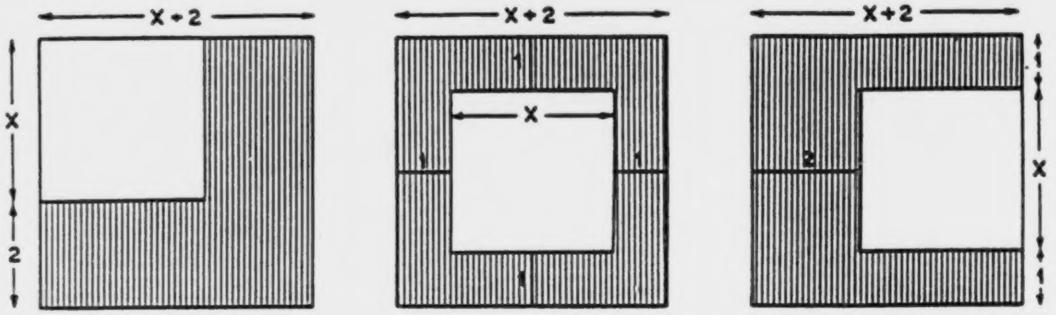


Figure 2 : Some possibilities for the extended square

An open phrase is obtained for the additional area (4x+4, x represents a positive integer) with which one can "play" and obtain some general conclusions about image sets for certain subsets of the domain. In the next stage we start with an open phrase for an additional area 5(2x+5), and work back to the situation from which it arose, with a hint to the two variable situation which comes later. Here is another opportunity to try to build generalizations and to test them, this time the added length is an odd number (5). In the third part there is an open phrase in two variables b(2x+b); less familiar and more demanding on the student exploring this case. This is the end of the version for the average students (Squares, level I).

We prepared an experimental version of the project and, as usual, tested it in teachers' workshops. From the work and discussion with the teachers we gained a clear impression that it was good material for an assignment project, but would be too easy for more able students and too difficult for the weaker ones. Thus we reworked the experimental version, on the one hand for the weaker students by adding auxiliary sections, and on the other hand, we enlarged the project for able students.

In the version for the more able students (Squares, level II), the stages described above are reduced and the project is enlarged further. The restriction to integers is removed and new problems on the relation between changing sides and area of squares arise: the sides decrease, the side is fixed and the extension is a variable, the extension is proportional to the side, one pair of sides increases by b and the other decreases by b.

There is almost no limit to the imagination and ideas that can be developed, but does the working on such projects really enhance student mathematical activity? Do such activities elicit the desired goals? For answers to these questions we have to analyze and evaluate student work.

4. INVESTIGATION RELATED TO GOALS

4.1. Descriptive analysis

The learning goals and the material were "new" to teachers. In addition to out-of-school teacher workshops, there was a need for in-school guidance, to "optimize" the use of the material. While providing the teacher guidance, we also had a chance to observe student work on the projects.

In 1978 an investigation study was conducted. The population which worked on the assignment project *Squares* (see Appendix A) was eight 9th grade classes (age 14-15). Four classes (n=81) were given the easier version *(Square I)*, and four (n=131) the harder version *(Square II)*.

The students were asked to do the first sections of each project at home without help and to bring their attempts to the teacher. Their work was checked and annotated by the teacher with our assistance. Then a lesson was devoted to discussing the project, in which students were invited to present their work, to cooperate with other students and ask for advice. They were then, asked to improve their responses and to proceed at home. There were about three to four such stages for each project. After a certain period the students gave their projects to the teacher for final checking. Some descriptive analysis of the investigation of student work related to the goals will be given here.

The first sections are very simple and were inserted to lead into the project. Most students managed to get started by themselves. This is significant since students tend to avoid unfamiliar tasks for a variety of reasons. Sometimes they are convinced, without trying, that it cannot be done, or they use the unfamiliarity as an excuse for giving up. On the other hand, they may want to try but hesitate or do not know how to approach the problem. Hence it is important that the first items are, if slightly unfamiliar, very simple. We then gradually progress to the unfamiliar and not so simple. For example, section III(c) in *Squares* (level I):

Given that the additional area is 175 cm^2*, what can you say about the side length of the original square?*

The related equation is $b(2x+b) = 175$ where x denotes the side length and b denotes the added length. Following part II of the project, students wrote 5 for b and got $x = 15$. Some were not satisfied with the quality of answer and an argument we heard went as follows:

- We need to find a value for b and a value for x.
- Perhaps there is more than one value. The question says

"what can you say..."

- We had that in the previous section and, yet, we found just one number.
- That doesn't mean that the same thing will happen here...

We can assume that for the students involved, the *quality of answer* goal was achieved.

Only few obtained the full solution:

b must divide 175, thus b can be either 1, 5, 7 or products of these.

for $b = 1$ $\quad 2x + 1 = 175 \quad$ *thus* $\quad x = 87$

for $b = 5$ $\quad 2x + 5 = 35 \quad$ *thus* $\quad x = 15$

for $b = 7$ $\quad 2x + 7 = 25 \quad$ *thus* $\quad x = 9$

There is no point in going further, since $x > 0$ *and hence* $2x + b > b$.

Another example, which illustrates the *quality of procedure,* is that provided by *Squares* (level II), section I(f). In this case, each side of a square of length x is extended by 5 and the question is:

Is it possible to have a situation in which the additional area is an integer multiple of the area of the original square?

Many students rephrased the problem in a mathematical form equivalent to the equation

$$kx^2 = 10x + 25,$$

k an integer, for which we have to find integer values for x.

Most of the students used a non-systematic guess and check procedure and some found the only two possibilities x = 1 and x = 5. Yet others attempted systematic checking and of these, few succeeded by this approach unaided. The remainder completed after receiving guidance. The guidance took the following form

$$k = \frac{10x+25}{x^2}$$

Let's try to guess possible values for x.

How many values are we going to need to check? Are we going to live long enough? (Since 10x+25 is an odd number, x^2 *and therefore x must be also odd.)*

What happens to k as x increases? k decreases.

When is $k < 1$ *? (*$x > 13$*, so all we need check is* $1 < x < 11$*, x odd.)*

This is an example of interaction of teacher and students towards a better *quality of solution.*

Some general unaided attempts at the problem were also observed. For example, an incomplete argument:

$$kx^2 = 10x + 25$$

$$k = \frac{10}{x} + \frac{25}{x^2}$$

x^2 *must divide 25* $\Rightarrow$ x^2 *is 1, 5, or 25*

x *must divide 10* $\Rightarrow$ x *is 1, 2, 5 or 10.*

Note - This response is incomplete, since one has to show that the sum cannot be an integer unless both fractional expressions are also integers.

Two elegant and original arguments, which can be considered as extension of procedure, were given:

$$k = \frac{(x+5)^2 - x^2}{x^2}$$

$$k = (1 + \frac{5}{x})^2 - 1$$

k is an integer for $x = 1$ *or* $x = 5$ *only.*

$$kx^2 = 10x + 25$$

$$x(kx-10) = 5 \cdot 5$$

Thus x can be only 1 or 5 (or 25 which fails).

Section I(g) in *Squares* (level II) was meant to surprise. In the previous sections the students noticed that the additional area increases with increasing x. Here the percentage ***decreases*** with increasing x. Several students who had previous experience on other projects, on their own initiative, with no leading question, remarked that the percentage decreases with increasing x *(extension of answer)*. In the following section (I(h))the student has to find the side of the square for which the increase in area is closest to 100%. One can find the answer by guessing and it is always better to do this intelligently, i.e. by using the result of section g *(quality of procedure)*. Several students remarked that they found the answer (12) after a lot of trials, others constructed an organized table. One boy came up with the following: "I found that for x = 12, the percentage is close to 100%. I wanted to know for what x it is exactly 100%. My brother told me that I need to solve a quadratic equation and I have not learnt that yet. Isn't there another way?" *(extension of question,* unprompted).

Fortunately one of the girls in the class answered section f using

$$k = \left(1 + \frac{5}{x}\right)^2 - 1 .$$

We asked her to write the expression on the blackboard and asked the boy if he could answer his question now. He could!

As a final word, we quote one student, who did a very nice piece of work on *Squares* (level II). When he reached part V he was tired and wrote, "I will not do this part because I don't feel like doing it and I don't care what mark I get. Not because I don't know how to do it, just I don't want to. But I'll do part VI (the open question) because here I've several nice ideas."

We annotated this: "In a project like this, you don't have to do everything, and we agree that it is much more fun to work on the bits where one has several nice ideas ... So why not?"

4.2. Qualitative analysis

In an assignment project there are many sections, only some of which are challenging and intended to be directly related to the goals (non avoidance, quality of solution, and extension).

We illustrate the analysis of student responses in terms of the goals, by looking at the responses to such a section (I(g))in *Squares* (level I) before the students received any guidance.

In the preceding sections (I(e), (f), the student found that the expression for the additional area is $4x + 4$ or $4(x + 1)$. He was led to conclude that the largest common factor, whatever the additional area, is 4 when x is even and 8 when x is odd. Now he has to find the first four integer values of x for which the additional area is divisible by 16 and to establish a pattern.

Only 2% avoided this section.

Three types of correct answer were obtained:

A_1 - numerical answer:

the four numbers 3, 7, 11, 15 without any further attempt to establish a pattern;

A_2 - "finite" pattern:

the four numbers and an "internal" pattern, e.g. "There is a difference of four between the numbers."

A_3 - "infinite pattern"

"every second odd number starting from 3".

Three types of procedure in the construction and testing of the answer were identified:

P_1 - numerical substitution;

P_2 - incomplete reasoning, most of the time verbal;

P_3 - complete formal proofs based on an equation such as

$4x + 4 = 16k$.

These procedure types were found in combinations with the various types of answers. Responses that extended the task occured, if rarely. But the few we did receive were interesting.

(i) Further investigation for cases in which the additional area is divisible by 32.

(ii) A more general investigation, involving classes of natural numbers: in every case the additional area is divisible by 4. If we restrict ourselves to odd numbers, then their substitution in the expression will give numbers which are divisible by 8. If we continue this way and restrict ourselves to every second odd number, the result will be divisible by 16 and so on - divisible by 32, 64, 128 ... The process has no end.

To assess the quality of solution, each response is classified on two scales: answer and procedure.

In the example given above, three levels of quality of answer and three levels of quality of procedure were identified.

Quality-of-answer scale (A):

1. primitive, 2. partial, 3. full.

Quality-of-procedure scale

1. trial, 2. incomplete argument, 3. complete argument.

These are specific to each section, but have general characteristics which can be applied to any section relevant to our goals.

The quality of solution is considered as a combination of answer and procedure, and each student response can be allocated to a cell in the following table.

Answer (A) Procedure (P)	A_1 Primitive	A_2 Partial	A_3 Full
P_1 Trial			
P_2 Incomplete argument			
P_3 Complete argument			

Table 1 : Quality of solution table

The extension goals, since by their nature they occur in a minority of responses only, were considered separately.

The evaluation framework provides a measure of progress.

Improvement of mathematical activity will be exhibited in the *quality of solution* table as a "population shift" in a "south and/or easterly" direction.

To summarize, the outcome of the investigation study was a framework for analysing student work on the assignment projects in terms of the goals.

5. EVALUATION INSTRUMENTS

5.1. General information

The analysis we have described can be used to evaluate student work on the Assignment Projects, but by itself does not distinguish between the mathematical maturity achieved as a result of regular curriculum studies, and the expected special contribution of working on the projects. If we want to evaluate the effectiveness of the treatment we need a relevant research instrument, by which we can assess mathematical activity. Such an instrument can be used to measure progress of students who work on projects, and also to compare them to students who have not had this experience. Based on the same goal-framework as described for the treatment, an *assignment test* (Zehavi and Bruckheimer, 1982a) was developed. In addition to the *assignment test* (cf. 5.2.), two standard instruments were prepared, an *achievement test* (cf. 5.3.) and an *attitude questionnaire* (cf. 5.4.). The achievement test (ACH) contains 20 multiple choice items on some of the topics of the ninth grade curriculum in algebra. It is designed to provide information about student success in regular mathematics studies. The attitude questionnaire is Likert type. It contains two sub-questionnaires, (ATSI) student self image and (ATMA) student image of mathematical activity.

5.2. Achievement test

A multiple-choice achievement test (ACH) containing 20 items was prepared. The test deals with some topics in the ninth grade algebra curriculum on the application of general function concept (the test is included in Appendix B). It was designed to provide information about student success in the regular mathematics program. This information was needed to

a) choose matching population samples for the experimental and control groups in the study;

b) investigate whether working the assignment projects affects achievement in the regular mathematics programme.

A first version was applied in the preliminary comparative study in 14 classes (n=351) and was modified as a result of item analysis. It was decided to include in the test items which have high discrimination index (about 0.5), with difficulty index ranging from 0.30 to 0.80. The final version was piloted in 9 classes (n=235) and was found to have a mean score of 57 with standard deviation of 22. It's α-Cronbach reliability index was $\alpha = 0.82$.

The test contains items at the various cognitive levels, as assessed by experts and by the teachers of the classes in which the test was applied.

The teachers were further involved in the analysis of the achievement test. They were asked to predict the success rate of their students on each item and their predictions were compared with the actual performance. In general consistent mismatch was found for topics which were related to "modern" mathematics and "new" instructional methods. (Detailed description of the method of analyzing tests using the teacher's predictions and some of its applications can be found in Appendix F.). These findings indicated that a lot of thought and attention was needed in the teacher implementation of the assignment projects.

5.3. Attitude questionnaire

A Likert type attitude questionnaire was prepared in two parts. The first part was designed to provide information about student self-image in mathematics (ATSI), and the second part about student image of mathematical activity (ATMA). While the ATSI is usual in attitude questionnaires, the ATMA is more specific to this study.

The items in the two subquestionnaires were mixed and presented as one attitude questionnaire (see Appendix B). The role of the attitude questionnaire in the study was to investigate whether working assignment projects affects student attitudes to mathematical activity.

A first version with 48 items was prepared, containing 24 items belonging to each subquestionnaire. The classification was verified by experts. The students were expected to choose their response from one of the following: *strongly agree, agree, neutral, disagree, strongly disagree.*

This version was applied in the preliminary comparative study (n=351). The reliability of the whole questionnaire was calculated, $\alpha = 0.70$. Many students choose the *neutral* option and thus it was decided to use only two options in the future: *tend to agree, tend to disagree*. The students were also asked to specify unclear and annoying items. Their comments were taken into consideration together with item analysis findings and the questionnaire was modified and shortened. The new version contained 32 items only. It was piloted (n=235) and was found to have a higher reliability $\alpha = 0.81$. The reliability of the ATSI (15 items) was 0.88 and that of the ATMA (17 items) 0.75. In order to sharpen the instrument further, it was decided to omit all "obvious" items, with which almost everybody agreed (or disagreed). This left 10 items in the ATSI and 12 items in the ATMA.

5.4. Preliminary comparative study (1979)

One of the objectives of the field studies at the preliminary evaluation stage, was to obtain some comparison between students who worked on the assignment projects and students who did not. Six top stream ninth grade classes (n=169) participated in the experimental group and eight classes (n=182) in the control group. The students in the experimental group worked on three or four projects over a period of four months. Both groups were given the unmodified version of the achievement test as pretest and posttest. There was a significant improvement (as measured by the t-test) in the experimental group and insignificant improvement in the control group. At the posttest stage, the first version of the attitude test was also applied, showing non-significant difference between the two groups. At that time the assignment test was not yet available. Thus, in order to see if the treatment had any goal-related effect, an open question, in the form of a sort of "mini" project was given.

In this question, the students had to tackle a problem which, at first sight, seemed to be on something they had not yet come across; they had to establish a conjecture and prove it. The text of the question is as follows:

$\frac{a + b}{2}$ *is the arithmetic mean of two positive numbers a and b.*

$\sqrt{ab}$ *is the geometric mean of two positive numbers a and b.*

Of the following four statements, which one is correct?

(a) The arithmetic mean is always equal to the geometric mean.

(b) The geometric mean is always greater than or equal to (but never less than) the arithmetic mean.

(c) The arithmetic mean is always greater than or equal to (but never less than) the geometric mean.

(d) Sometimes the geometric mean is greater and sometimes the arithmetic mean is greater.

Explain the reason for your choice of the correct statement. (At the bottom of the page you will find six open sentences. If you wish, you can construct a proof for the correct statement by ordering these appropriately and adding an explanation for each step.)

1. $(a+b)^2 > 4ab$
2. $\frac{a+b}{2} > \sqrt{ab}$
3. $a^2 + 2ab + b^2 > 4ab$
4. $(a-b)^2 > 0$
5. $a + b > 2\sqrt{ab}$
6. $a^2 - 2ab + b^2 > 0$.

In the experimental group 81% selected the correct option and 90% of the group gave some reason for the answer they chose. In the control group only 60% chose the correct option and 56% added some reason. But what is more interesting, in the experimental group there were significantly more attempts to start a proof (45% gave at least a partial proof and 10% a completed proof). In the control group only 22% attempted a proof and only 3% fully succeeded.

There was here some clear first evidence that work on the projects had a cumulative and beneficial effect on students' approach to, and achievement in, mathematical activity.

Before we could continue to the main study we had to have the main research instrument, the *assignment test*.

5.5. Assignment test

5.5.1. Design and construction

The test contains four tasks and can be applied during one class period. Three items are in fact, mini-assignment-projects, in the sense that they have, to some degree, elements related to the goal-framework. They were arranged according to difficulty previously determined. The last item is different from the rest and specifically asks for extension.

The experience we had with the assignment projects helped in the construction of several possible items. The appropriateness of these items was judged by a group of 8 experts, who first did a theoretical task analysis. Then, eight project items were given to students (n=130), and the responses were analysed by the experts to see if they can be assessed in quality of solution tables (Table 1). The most appropriate items were included in the pre- and post-test (given below). Since the research was designed to have experimental and control groups, it was decided to use one item in both, the pre- and post-test, in order to sharpen the pre-post comparison.

Thus assignment 2 appears in both versions with only a "technical" modification to avoid rote recall.

As for assignments 1 and 3, we had a general impression that those of the pretest and those of the posttest are rather similar and could be matched in the analysis process. It will be confirmed in a further stage (cf. 7.3).

Assignment Test (Pretest)

Assignment 1 : Which numbers when substituted in $3x + 1$ will give even numbers?

Assignment 2 : Given the expressions $x^2 - y^2$ and $(x - y)^2$. Find pairs (x, y) which when substituted in the above expressions give the same result.

Assignment 3 : Given the equation $b(2x + b) = 175$, where x and b represent positive integers; what can you say about x and b ?

- - - - - - - - - - - - - - - - - - - -

Assignment 4 : Pose and try to solve question(s) similar to those which appear above.

Assignment Test (Post-test)

Assignment 1 : Which numbers when substituted in $x^2 - 5x$ will give negative numbers?

Assignment 2 : Given the expressions $y^2 - x^2$ and $(y - x)^2$. Find pairs (x, y) which when substituted in the above expressions give the same result.

Assignment 3 : Given the open sentences $50 + 2a + b = 7b$ and $50 + 2a + b < 100$, where a and b represent positive integers, what can you say about a and b?

- - - - - - - - - - - - - - - - - - - -

Assignment 4 : Pose and try to solve question(s) similar to those which appear above.

Evaluation Schemes

The guiding principles for the evaluation scheme were that

1) it should reflect the goals of the projects;
2) it should be suitable for the analysis of a large number of responses;
3) it should be as clear and unambiguous as possible to give valid (objective) assessment.

The group of experts was divided into work teams and they prepared evaluation schemes. For each assignment, they prepared a detailed scheme and decided on the three main categories of the quality-of-answer and quality-of-procedure scales.

As an example of an evaluation scheme, we bring here the one devised for the assignment which appears in both versions of the test (Zehavi and Bruckheimer, 1984). The evaluation schemes for the other assignments are given in Appendix B.

The quality of solution score (Q) is calculated as the sum $Q = i + j$, $(i,j = 1,2,3)$ for a A_i/P_j response. Thus the maximum value of Q is 6. If a response is very weak and cannot even be considered as A_1/P_1, the evaluator assigns $Q = 1$. An example of such a response on assignment 2 is the substitution of two *different* pairs of x, y in the two expressions $x^2 - y^2$ and $(x - y)^2$. No attempt to answer is assigned $Q = 0$.

Main categories and refined evaluation scheme for assignment 2

Answer scale

A_1 (at most) numerical examples for $x = y$ only or $y = 0$ only

11 all (x, y),
12 the empty set,
13 incorrect example, e.g. $(0, 1)$ $(5, -5)$,
14 examples of $x = y$,
15 examples of $y = 0$,
16 the pair $(0, 0)$,
17 other.

A_2 no further than the generalization $\{(x, y)/\ x = y\}$ only, or $\{(x, y)/\ y = 0\}$ only

21 numerical examples of $x = y$ and of $y = 0$,
22 generalization $x = y$,
23 (incorrect) generalization $|x| = |y|$
24 (incorrect) x any number and $y = 0$, or $x = 0$ and y any number,
25 (incorrect) $x = 0$ and y any number,
26 other

A_3 complete generalization $\{(x, y)\quad x = y \text{ or } y = 0\}$

31 (incorrect) generalization $x = y$, $x = 0$, $y = 0$,
32 generalization $x = y$ or $y = 0$,
33 other.

Procedure scale

P_1 checking numerical examples or algebraic manipulation of expressions

11 no explanation,
12 algebraic manipulation of expressions,
13 checking numerical examples,
14 the equation $(x - y)^2 = x^2 - y^2$, and no more,
15 other.

P_2 proof of sufficiency

(incomplete proof of sufficiency)
21 algebraic and logical mistakes,
22 algebraic mistakes,
23 logical mistakes,
24 $x = y$, the result is zero,
25 $y = 0$, both results are x^2,
26 24 and 25,
27 other.

P_3 proof of sufficiency and necessity

31

$$\begin{aligned}(x - y)^2 &= x^2 - y^2\\ x^2 - 2xy + y^2 &= x^2 - y^2\\ 2y^2 &= 2xy \ /:2y\\ x &= y\end{aligned}$$

32

$$\begin{aligned}(x - y)(x - y) &= (x + y)(x - y) /:(x - y)\\ x - y &= x + y\\ y &= 0\end{aligned}$$

33 31 and 32 separately,
34 finding the complete truth set $x = y$ and $y = 0$ (correctly),
35 other.

5.5.2. Validation studies

The content validity of the assignment test was very carefully considered in the process of its construction (theoretically and empirically, as described above). After the test was prepared a study was conducted to verify (a) its construct validity within the framework (Cronbach, 1960) and (b) the reliability of the evaluation scheme.

(a) Construct validity

Based on the findings of the preliminary comparative study the assumption was that work on the projects does elicit the goals. The test and its evaluation scheme were prepared within the goal-framework. Now we wanted to find out whether it is "sensitive" and can distinguish between groups who worked or did not work on projects.

Two 9th grade classes were given the multiple-choice achievement test (ACH) and then one class worked for about one month on the project $n^2 - 1$ (Bruckheimer and Hershkowitz, 1977). The average ACH score of the project-class (n=20) was 63 and the average ACH score of the non-project class (n=29) was 68 (the difference is not significant). After this month the pretest version of the assignment test (which was needed for the beginning of the main study), was given to both classes and the results compared. In Tables 2 and 3 we see the quality of solution cumulative distribution for assignments 1-2-3 in each class. We can see that the project class performed considerably better, which may support the validity of the instrument.

Answer (A) Procedure (P)	A_1 Primitive	A_2 Partial	A_3 Full	total
P_1 Trial and error	11.7	10.0	1.7	23.4
P_2 Incomplete argument	11.7	27.1	6.6	45.4
P_3 Complete argument		20.0	10.0	30.0
total	23.4	57.1	18.3	98.8

Table 2 : Quality of solution distribution (in %) for assignments 1-2-3 in the project-class (n=20)

Answer (A) Procedure (P)	A_1 Primitive	A_2 Partial	A_3 Full	total
P_1 Trial and error	18.4	12.6	2.3	33.3
P_2 Incomplete argument	19.5	31.1	1.1	51.7
P_3 Complete argument		9.2	2.3	11.5
total	37.9	52.9	5.7	96.5

Table 3 : Quality of solution distribution (in %) for assignments 1-2-3 in the non-project-class (n=29)

(b) Reliability of evaluation scheme

In order to check the reliability of the evaluation scheme, 12 students (6 from each class) were sampled randomly from three groups, based on the scores on the achievement test (2 low, 2 medium, 2 high). Their responses were given to five teachers to assess the quality of solution according to the scheme. They assigned each response to one of the categories in the detailed evaluation scheme, both on the answer scale and on the procedure scale.

Table 4 shows the results of the five teachers (a - e) for assignment 2 for the 12 student sample.

	Teacher	a			b			c			d			e		
	Student	A	P	Q	A	P	Q	A	P	Q	A	P	Q	A	P	Q
project class	1	11	12	2	11	12	2	11	11	2	17	12	2	11	12	2
	2	32	34	6	32	34	6	32	34	6	32	35	6	32	34	6
	3	32	34	6	32	34	6	32	34	6	32	34	6	32	34	6
	4			1			1	1	1	2			1			1
	5			1			1			1			1			1
	6	22	13	3	14	13	2	14	13	2	14	13	2	14	13	2
non-project class	7			1			1			0			1			1
	8	22	13	3	22	13	3	22	13	3	14	15	2	22	15	3
	9	22	24	4	22	24	4	22	13	3	22	24	4	22	24	4
	10	24	13	3	24	21	4	24	21	4	24	13	3	24	15	3
	11			1			1			1			1			1
	12			0			0			0			0			0

Table 4 : Evaluation by five teachers of student responses to assignment 2

As can be seen, reasonable agreement between the five teachers was obtained, and similar results were obtained for the other assignment test items.

Further confirmation of the reliability of the evaluation scheme on a larger scale was obtained in the communication study (see 7.3)

After the evaluation scheme was established, we considered the statistical aspects of the research instrument.

The main study was designed for hundreds of students.

We expected progress and there was a need to be able to decide whether that progress is statistically significant.

5.5.3. Statistical test

For scores which are on an ordinal scale, nonparametric techniques of hypothesis testing are suited.

For pre-post comparisons of the Assignment Test, the Wilcoxon matched-pairs signed ranks test (Siegel, 1956) can be considered, since the test utilizes information of the magnitude, as well as the direction of the differences within pairs.

We have to rank the differences and to find T, the smaller sum of like-signed ranks. For large samples, T is practically normally distributed with a mean

$$\mu_T = \frac{n(n+1)}{4}$$

and standard deviation

$$\sigma_T = \sqrt{\frac{1}{24}\, n(n + 1)(2n + 1)} \quad .$$

Therefore,

$$u = \frac{T - \mu_T}{\sigma_T} = \frac{T - \frac{n(n + 1)}{4}}{\sqrt{\frac{1}{24}\, n(n + 1)(2n + 1)}}$$

for $n > 60$, is approximately normally distributed with zero mean and unit variance.

In using the Wilcoxon signed-rank test, the ranks corresponding to zero differences are deleted. In the case of large samples and few values for the variables, many ties occur, both non-zero ties and zero ties. Cureton (1967) suggested a modification of the formula for normal approximation for large samples when zero differences are present. According to his version, one considers the zero-differences when assigning ranks, but not when computing the rank sum T. We have,

$$\mu_T = \frac{n(n + 1) - t_o(t_o + 1)}{4}$$

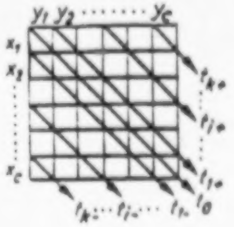
$$\sigma_T = \sqrt{\frac{1}{24}\left[n(n + 1)(2n + 1) - t_o(t_o + 1)(2t_o + 1)\right] - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{48}}$$

t_o - number of zero ties,

t_i - number of ties with absolute difference i,

k - the largest value of the absolute difference.

Raatz (1977) suggested an efficient algorithm for performing the test. First we organise the data in a cross table so we can see the pairs:


y_1 y_2 .... y_c
x_1
x_2
x_c
t_{k+}
t_{i+}
t_{1+}
t_0
t_{k-} ... t_{i-} ... t_{1-}

Then we complete the following table:

	1	2	3	4	5	6
	+ t_i	−	t_i	t_{ic}	t'_{ic}	t_i^3
0	▨	▨			▨	▨
1 2 ⋮ k						
	$S^+ = \ldots$	$S^- = \ldots,$				$A = \ldots$

$$t_i = t_{i+} + t_{i-}$$

$$t_{ic} = \sum_{j=1}^{i} t_j$$

$$t'_{ic} = t_{ic} - \frac{1}{2}(t_i - 1)$$

T is calculated as the product of the column (1) or (2), whichever is the minimum value (S^+, S^-) and the column (5) of t'_{ic}.

The expected T is,

$$E(T) = \frac{n(n+1) - t_0(t_0+1)}{4}$$

and the variance

$$V(T) = \frac{1}{24}\left[n(n+1)(2n+1) - t_0(t_0+1)(2t_0+1) - \frac{1}{2}(A-S^+-S^-)\right]$$

Thus for the Wilcoxon-Raatz test for large samples, one needs to check the significance of the normally distributed U

$$U = \frac{T - E(T)}{\sqrt{V(T)}} .$$

Lienert (1978) brings as example an application of this statistical test to evaluate the efficiency of smoke-stop treatment. A group of smokers (n=100) is given ranks before and after the treatment according to the amount of cigarettes each subject smokes per day.

This method seems applicable to our need in the comparative evaluation study (see 6.3.). We can assign a Q rank for each assignment and match the ranks of the corresponding assignments in the pretest and posttest.

6. COMPARATIVE EVALUATION STUDY

6.1. The study (1980)

The population was again top stream ninth grade classes. We started with 11 classes in the experimental group and 30 classes in the control group. The comparatively large number of control classes was sampled in order to enable a reasonable match of the two groups. The matching was based on the achievement test, which was given to all the classes.

Classes for which the average score was less than 40 were excluded, since our experience indicated that most students in such classes could not work on the projects at a high cognitive level.

In the experimental group the highest class average score on the achievement test was 71.3 and in the control group, there were some classes with much higher average score. Classes whose average score was above 75 were also excluded. The remaining classes (8 in each group) were then given the assignment (pre) test and the attitude questionnaire.

The students in the experimental classes worked on 3 - 4 projects. They were asked to do the first part of each project at home and bring their attempts to the teacher. Their work was checked and annotated by the teacher with our considerable help. Then a lesson was devoted to discussing the project, upon which the students were asked to proceed and invited to get advice and help as needed. After a given period they presented the project for final checking.

The researcher and members of the mathematics group in the Weizmann Institute were actively engaged in the implementation of the projects in the classes. Most of the written and verbal guidance to students was given by them. This intervention had two related reasons. Teachers had difficulty with this type of work, which was new to them. Also, in a research study of the effectiveness of the material we wanted to optimize its implementation. Consequently, all the experimental classes received almost the same treatment and can be considered as one group for the analysis of results. After about four months, the posttest version of the assignment test, the achievement test and attitude questionnaire were given. In the experimental group we had complete data for 191 students and in the control group for 194 students.

The achievement test and attitude questionnaire were multiple-choice and thus checked by the computer. The assignment tests were evaluated by the researcher according to the schemes described in section 5.2.1. Random samples were double checked by colleagues, who were not told whether they belonged to the experimental or control group. This procedure was followed to (a) revalidate the evaluation technique and (b) to ensure unbiased checking.

Appendix C contains information about group comparison of average scores and distribution tables. For almost all the parameters there was no statistically significant difference at the pretest level, but significant differences in favor of the experimental group were found at the posttest level. The major findings will be presented and discussed in the following two sections.

6.2. General findings

Table 5 presents the average scores for the two groups.

Legend:

ACH - achievement test (maximum score 100)

ATSI - attitudes: self-image (maximum score 100)

ATMA - attitudes: mathematical activity (maximum score 100)

Q - assignment test (maximum score 18, calculated as sum of Q_i

$i = 1, 2, 3$, for the three assignments

Test Group		ACH	ATSI	ATMA	Q
Experimental n = 191	PRE	59.2	68.2	67.8	8.8
	POST	66.6	67.2	68.4	10.5
	DIF	7.4	-1.0	0.6	1.7
	t	6.21	-1.16	0.72	6.90
	p	0.0001	n.s.	n.s.	0.0001
Control n=194	PRE	57.4	67.0	67.9	8.7
	POST	58.1	64.6	65.4	8.1
	DIF	0.7	-2.4	-2.5	-0.6
	t	0.64	-3.13	-3.36	-2.74
	p	n.s.	0.002	0.001	0.006

Table 5 : Pre-post comparison of average scores

Discussion

The same achievement test was given to both groups as pretest and posttest. Its first application was immediately after the subject matter was studied as part of the regular curriculum. The experimental group did significantly better on the test the second time round, whereas the control group showed no significant improvement. This may be explained by the fact, that students working on the projects need to review and use material from the curriculum, which affects their retention and understanding of these topics. Also the higher level of activity demanded by the projects, may affect their approach to "simpler" tasks.

To illustrate the above remarks, we give below the text and results for three items from the achievement test, for which there were significant differences between the two groups. In each case, the experimental group's scores increased from pretest to posttest and the control group scores decreased.

Item 2 *Given the function* $f : x \longmapsto 2x + 5$

the origin of -1 is

(a) -2 *(b)* 2

* *(c)* -3 *(d)* 3

The scores on this item were :

	pre	post
exp.	57	66
con.	53	43

Item 12 *The points (-3, 0), (2, 0), (0, 0) are the intersection with the x-axis of the function*

* *(a)* $y = x(x - 2)(x + 3)$ *(b)* $y = (x - 2)(x + 3)$

(c) $y = x(x + 2)(x - 3)$ *(d)* $y = (x + 2)(x - 3)$

The scores on this item were:

	pre	post
exp.	42	61
con.	48	45

Item 13 *Consider the family of functions* $y = 2x + b$, *where b can be any real number.*

Which of the following sentences is true?

Among the graphs of the functions

(a) there is pair which meet on the x-axis

(b) there is a pair which meet on the y-axis

(c) there is a graph parallel to the x-axis

* *(d) there is a graph which passes through the origin*

The scores on this item were :

	pre	post
exp.	46	53
con.	47	40

Item 2 is a technique question, included in the program, which the students studied just before the pretest was given, and the success was about the average success on the test as a whole. In the posttest, this item measured retention and/or improved understanding of the concept involved.

The two other examples are at higher cognitive levels. They contain some element of unfamiliarity, yet the concepts of intersection points with the axes and parameters are dealt with extensively in the program. Hence, the significantly improved success of the experimental group indicates that work on the assignment projects has a beneficial effect on achievement on non-standard questions.

In the responses to the attitude questionnaire, there was a significant "worsening" in the control group, which could be explained by end-of-year lassitude, while in the experimental group there was no significant change.

Achievement tests and attitude questionnaires are standard tools in research. The main research instrument, which is specific to this study is the assignment test and a significant difference in favor of the experimental group was found.

The score (Q) on the assignment test was found to be in high correlation at the 0.0001 level with all variables, (Table 6).

Test Group			ACH	ATSI	ATMA
Experimental	n = 191	PRE	0.35	0.43	0.32
		POST	0.53	0.42	0.38
Control	n = 194	PRE	0.34	0.40	0.29
		POST	0.49	0.38	0.32

Table 6 : Correlation table of Q scores with the other test scores

The figures in the table show that the correlation between the scores on the assignment test and the achievement test increased at the posttest stage. This can be explained by the fact that the achievement test was first applied immediately after the associated topics were studied, whereas at the end of the year, after a lapse of time, it made greater demands on the student. Thus it became "more similar" to the assignment test.

6.3. Findings on the assignment test

Table 5 shows an average increase of 1.7 in the Q score for the experimental group and an average decrease of 0.6 for the control group. Figure 3 gives the different distribution of DIFQ (which is the difference between the two scores) among the two groups.

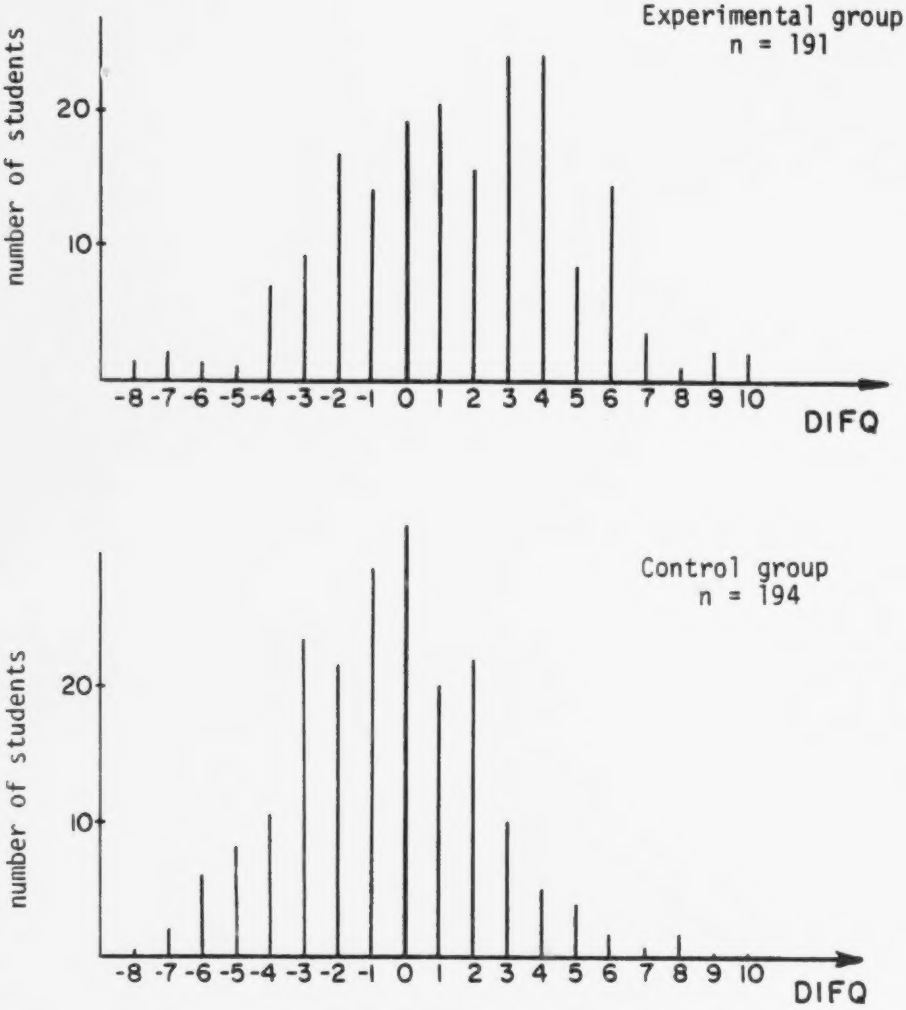


Figure 3 : DIFQ distribution

For the experimental group, we see a remarkable improvement for individuals, whereas for many students in the control group, the Q score has almost not changed.

Since the three assignments are of different difficulty levels, it is interesting to follow the changes for each of them separately and match the corresponding assignment in both applications.

We give below the results for both groups on assignment 2, which was identical in both the pretest and posttest. In Table 7 distribution in percentages of quality-of-solution, is presented. (The cells do not sum to 100%, since some students did not respond at all, or gave a solution that could not even be considered as A_1/P_1). Examples of student responses in the different cells are discussed in section 7.2.

Note: The results for assignments 1 and 3 are given in Appendix D.

experimental group

n = 191

PRE

P \ A	A_1	A_2	A_3
P_1	21.5%	33.5%	1%
P_2	5%	22.5%	1.5%
P_3		5.5%	0.5%

POST

P \ A	A_1	A_2	A_3
P_1	13.5%	20%	0.5%
P_2	3%	24%	2.5%
P_3		24%	5.5%

control group

n = 194

PRE

P \ A	A_1	A_2	A_3
P_1	17.5%	29.5%	2.5%
P_2	10.5%	22%	
P_3		10%	0.5%

POST

P \ A	A_1	A_2	A_3
P_1	22%	34.5%	
P_2	4%	18%	
P_3		12.5%	0.5%

Table 7 : Quality-of-solution distribution for assignment 2

Table 7 shows clearly the advantage of the experimental group. There is a population shift along the direction of the main diagonal. This becomes even clearer if we present the Q scores as in Figure 4.

It remains to answer the question whether the improvement is statistically significant. We shall use the Raatz method to represent the findings (Figure 5). This representation is appropriate since it demonstrates the actual "movement" from pretest level to posttest level. We use the Wilcoxon-Raatz test to check the significance of the "population shift". The results were highly significant for all three assignments (see Figure 5 and Appendix D), which show that material is effective.

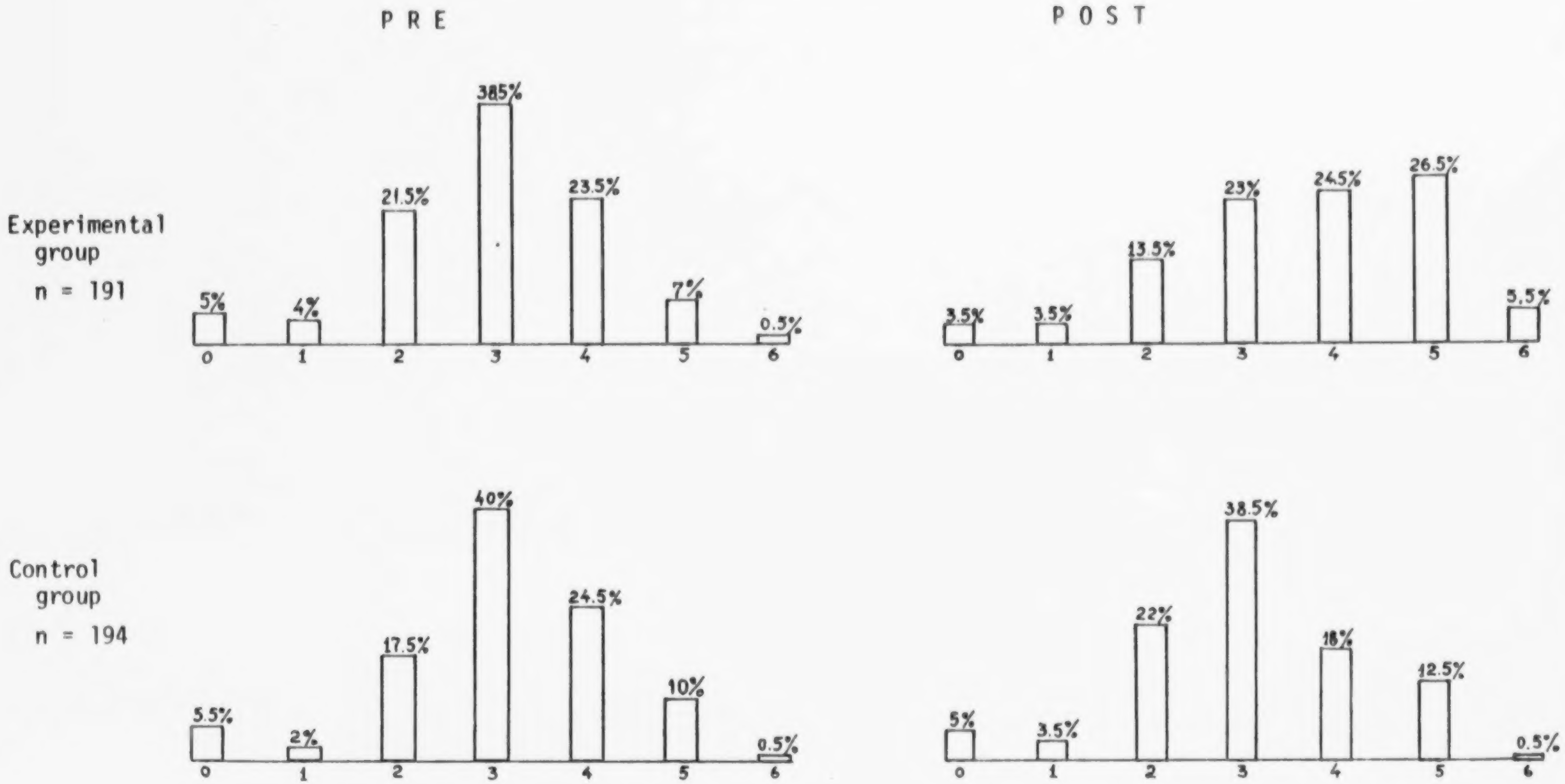
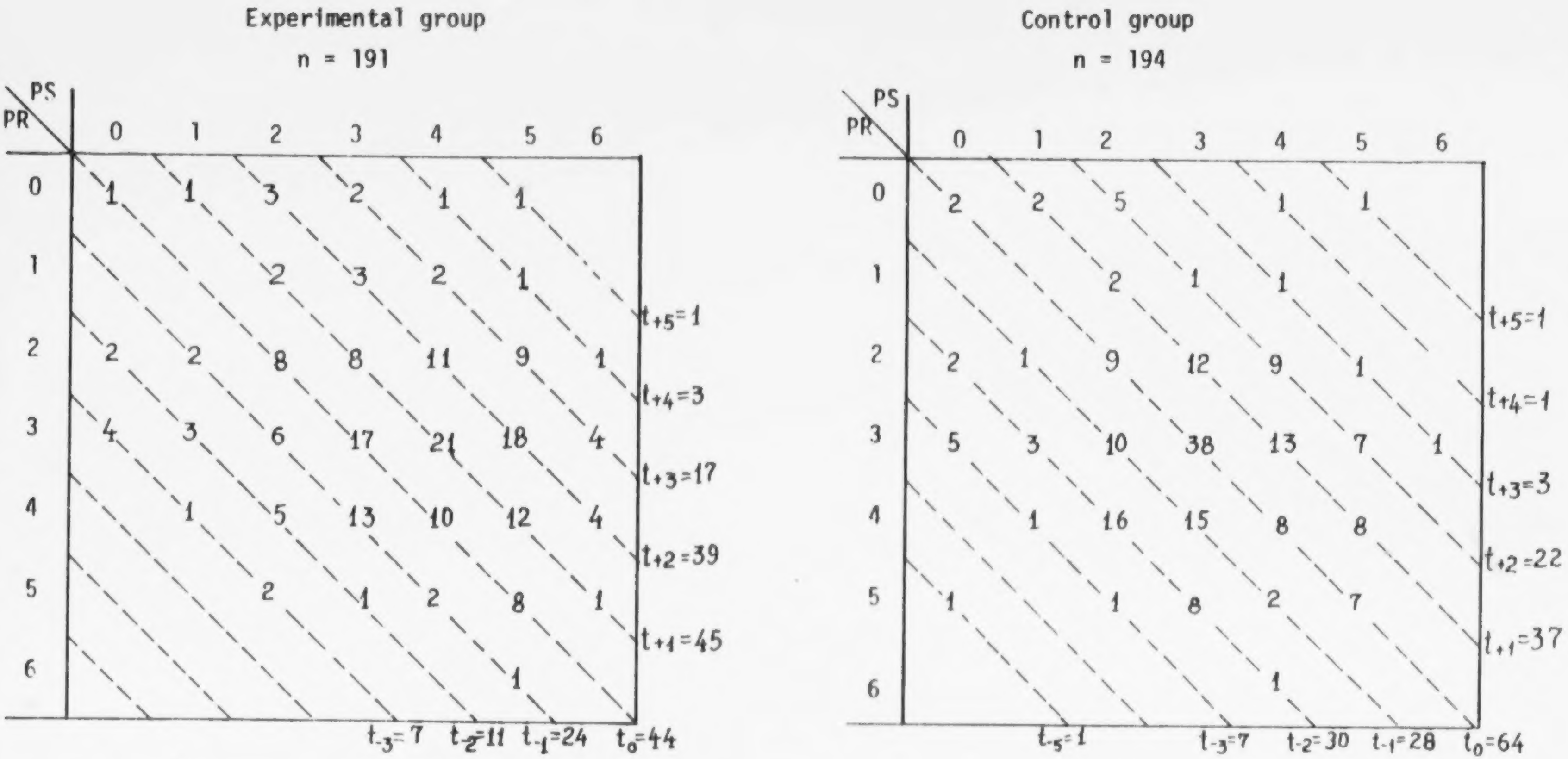


Figure 4 : Distribution of Q scores for assignment 2


Experimental group
n = 191
PS
PR
t+5=1
t+4=3
t+3=17
t+2=39
t+1=45
t-3=7
t-2=11
t-1=24
t0=44
Control group
n = 194
PS
PR
t+5=1
t+4=1
t+3=3
t+2=22
t+1=37
t-5=1
t-3=7
t-2=30
t-1=28
t0=64

Experimental group, n = 191

PR \ PS	0	1	2	3	4	5	6
0	1	1	3	2	1	1	
1			2	3	2	1	
2	2	2	8	8	11	9	1
3	4	3	6	17	21	18	4
4		1	5	13	10	12	4
5			2	1	2	8	1
6						1	

$t_{+5}=1$, $t_{+4}=3$, $t_{+3}=17$, $t_{+2}=39$, $t_{+1}=45$, $t_{-3}=7$, $t_{-2}=11$, $t_{-1}=24$, $t_0=44$

$S^+ = 105$ $S^- = 42$

T = 4648

E(T) = 8673

V(T) = 753.3

U = -5.339

one tailed P = 0.00001

Control group, n = 194

PR \ PS	0	1	2	3	4	5	6
0	2	2	5		1	1	
1			2	1	1		
2	2	1	9	12	9	1	
3	5	3	10	38	13	7	1
4		1	16	15	8	8	
5	1		1	8	2	7	
6					1		

$t_{+5}=1$, $t_{+4}=1$, $t_{+3}=3$, $t_{+2}=22$, $t_{+1}=37$, $t_{-5}=1$, $t_{-3}=7$, $t_{-2}=30$, $t_{-1}=28$, $t_0=64$

$S^+ = 64$ $S^- = 66$

T = 7955

E(T) = 8417.5

V(T) = 763.6

U = -0.605

one tailed p = n.s.

Figure 5 : Population shift for assignment 2
PR - pretest , PS - posttest

In section 3.1. we presented the goals of the assignment projects:

- *Non-avoidance of unfamiliar problems,*
- *Quality of solution,*
- *Extension of solution,*
- *Extension of question.*

The assignment test was designed to reflect these goals. Table 7 and Figures 4, 5 (and Appendix D) indicate a significant improvement of *quality-of-solution* on assignments 1-2-3. For the first goal, we look at the third assignment in each version which was meant to seem "unfamiliar". In the pretest application 10.5% of the students of the experimental group and 14% of the control group made no attempt (cf. Figure D5, Appendix D). In the parallel assignment in the posttest the same percentage of students in the experimental group avoided it, whereas the percentage of students (21%) in the control group increased significantly. The evidence is thus not very strong, and we cannot tell from our method of testing whether this goal was achieved. When students worked on an assignment project the "unfamiliar" sections were parts of the project. In such a situation it was easier to overcome this "block". The transfer to an unfamiliar item in the assignment test is not clear. Maybe if we used a 3-part assignment as a test with an "unfamiliar" third part the results would have been better.

As for the extension goals, we also did not obtain any decisive evidence from this application of assignment 4 ("pose and try to solve questions similar to those which appear above").

Many students in both groups, and in both applications of the test, only gave examples very similar to the given questions. There were some attempts by students of the experimental group to do something more interesting, but they had to stop when the period ended. Students who put considerable effort into assignment 1-3 did not have time for the fourth one. Here again, in the projects themselves, students posed and solved quite interesting problems, and may be we could get better results for such a task in a 4-part assignment.

However, some interesting *extensions of solution* were found in the experimental group at the posttest application.

We bring here some examples of extensions of assignment 2, whose solution is discussed extensively in sections 5.5.1. and 7.2.1.

For a start , one student commented

"You wanted to trap us into $(x - y)^2 = x^2 - y^2$!"

which in itself can be regarded as an extension of solution.

Another example of such an extension, is the following elegant deduction:

$$\begin{tikzcd}
 & x = y \arrow[dl] \arrow[dr] & \\
x - y = 0 \arrow[d] & & x^2 = y^2 \arrow[d] \\
(x - y)^2 = 0 & & x^2 - y^2 = 0
\end{tikzcd}$$

A rare, but suggestive phenomenon was visual induction by means of a graph. Having found several numerical pairs, these were plotted and found all to lie on a line.

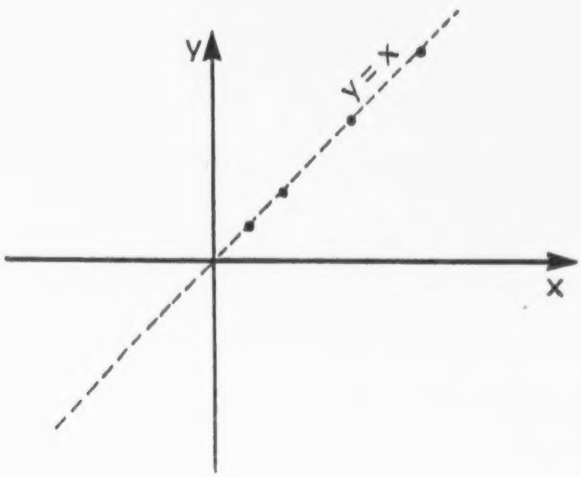


Figure 6: Induction by means of a graph

Many students reached the hypothesis $y = x$ or $y = 0$ by induction, but then realised that proof was needed to complete the argument. Worthy of special mention are four students (from three different classes) who followed this sequence, but their proof contained the cancellation flaw (*) :

$$\begin{aligned} &\text{"} \quad x^2 - y^2 = (x - y)^2 \\ &x^2 - y^2 = x^2 - 2xy + y^2 \\ &2xy = 2y^2 \\ (*) \quad &xy = y^2 \\ &x = y \text{ "} \end{aligned}$$

They expressed their dilemma in words of the form:
"How is it that the *proof* does not give me the *complete* solution?" Unable to answer, they escaped by drawing the graph of the truth set, as if to add conviction to the correctness of what they had failed to prove.

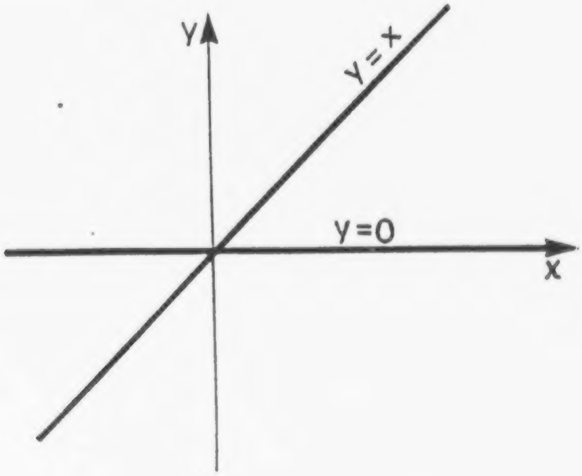


Figure 7 : Graphical conviction

7. DIAGNOSTIC EVALUATION OF MATHEMATICAL ACTIVITY

7.1. Post-evaluation studies

The comparative evaluation was done under experimental conditions. It took place in fact, in actual classes, but with strong intervention by experienced maths educators as explained in section 6.1.

The results of the comparative studies have shown that the assignment projects can indeed be effective in enhancing student mathematical activity. However, our experience indicated that students have difficulties in working on the projects. In order to guide the students to gradually overcome these difficulties, not only "good intuition" is needed, but also some knowledge of learning theories and psychology of learning.

If we want the material to be used effectively by teachers, it is very important to provide a practical method of evaluating student work on such "open problems" and hence, the means by which they can help students with suitable guidance.

We decided to achieve this aim by a deeper examination of student difficulties and a more detailed analysis of forms of effective guidance. The premise was, that if we could categorize the difficulties and the guidance into some sort of hierarchy, and establish a link between them, then we would have a model and a language which would help teachers in their work.

Up to this stage the studies involved fairly large population. To achieve our present aim, we felt it necessary to make a "sharper" application of the model in a "clinical" setting, to establish the guidance hierarchy for which we were searching.

This was followed by a "communication" study to examine whether the guidance hierarchy could be practically applied by teachers in the classroom.

Interpretation of the data from the main comparative study and the communication study in light of the diagnostic findings raised some more questions that need to be investigated in future studies.

7.2. Clinical study (Zehavi et al, in press)

7.2.1. Qualities of mathematical activity and guidance hierarchy

- A worksheet containing the three assignments from the pretest was given to a 9th grade class (n=30).

The following Answer/Procedure tables (Figures 8-10), in which the students are numbered from 1 to 30, give the distribution of unassisted student responses for the three assignments. (Students whose "solutions" did not reach even the A_1/P_1 stage, have been omitted.)

The diagnosis to this point is static and analyses the limit of student achievement. It does not provide us with insight into the reasons why a student ceased activity in a particular cell. To classify the obstacles that impede student progress, and the possibilities for remedial treatment, all the students were interviewed. The purpose of the interviews were three-fold:

(i) to confirm the validity of the classification;

(ii) to establish a helpful guidance system;

(iii) to diagnose difficulties.

The technique chosen for the interviews will be exemplified in the discussion of the hierarchy of solutions and hints.

A P	A_1 primitive	A_2 partial	A_3 full
P_1 trial		13 26 20 30	
P_2 incomplete argument		7 3 17 14 4 19 15 5 21 23 8 24 10 25 12 28 16 29	27
P_3 complete argument		2 9 11 18 22	1 6

Figure 8 : Answer (A)/ Procedure (P) table for assignment 1

A P	A_1 primitive	A_2 partial	A_3 full
P_1 trial	8 11 17	3 4 13 14 19 21	10
P_2 incomplete argument	20 28 25 26	30 7 9 12 23 27	5 22 24
P_3 complete argument		1 2 16 18 29	6

Figure 9 : Answer (A) / Procedure (P) table for assignment 2

A P	A_1 primitive	A_2 partial	A_3 full
P_1 trial	11 3 23 21 24 28 29 25	9 26	
P_2 incomplete argument	1 10 5 20 8 27	16 18 22	
P_3 complete argument			2 6

Figure 10 : Answer (A) / Procedure (P) table for assignment 3

Based on Skemp's ideas of different kinds of understanding (Skemp, 1979a, 1979b) we developed a hierarchy of hints.

Quality of hint	Immediate desired outcome and progress
(I) Instrumental	technique, simple operation
(IR) Instrumental-Relational	technique with simple relational progress (e.g. simple generalisation and/or establishment of pattern with some argument)
(R) Relational	e.g. establishment of pattern with justification
(RL) Relational-Logical	e.g. pattern with some proof
(L) Logical	proof

Table 8 : Hierarchy of hints

In the interviews, each of the students (except for the few whose responses on an assignment were classified A_3/P_3) was first given an L hint. In the many cases in which this produced no response, an RL hint was given and so on, with progressively lower-order hints, until some hint elicited progress. From the recording of the interviews, the level of the first effective hint was noted, as well as the extent of subsequent progress, i.e. the A/P cell now achieved by the student.

If the original response was in A_3/P_3, the student was asked to extend: i.e., to write his solution in formal "elegant" mathematics, and/or to pose another problem similar to the given assignment. We classified such a hint as *extension* (E). The application of this general scheme to the three specific assignments is discussed below. For each assignment we bring the Answer/Procedure categories, a sample of the hints used in the interviews and a table which summarizes hints and subsequent progress.

Assignment 1: *Which numbers when substituted in 3x + 1 will give even numbers?*

The main categories on the Answer scale were:

A_1 : some odd numbers;
A_2 : the pattern "all odd numbers:;
A_3 : all numbers of the form $\frac{2n - 1}{3}$ (n an integer).

The main categories on the Procedure scale were:

P_1 : guess and check procedure;
P_2 : proof of one condition (sufficient or necessary);
P_3 : sufficient and necessary proof.

The following are examples of hints used:

I - What do you get if you substitute for x#3, 4, $4\frac{1}{3}$, 5?
IR - What happens if you substitute for x#4872, 14357, 2/3, 3/4?
R - What happens if you substitute integers? What happens if you substitute non-integers?
RL - Suppose you substitute any other number (other than in your answer), couldn't you get an even outcome?
L - Are you sure that you have found all possible answers?

Figure 11 shows the type of the first effective hint in each case, and the subsequent progress.

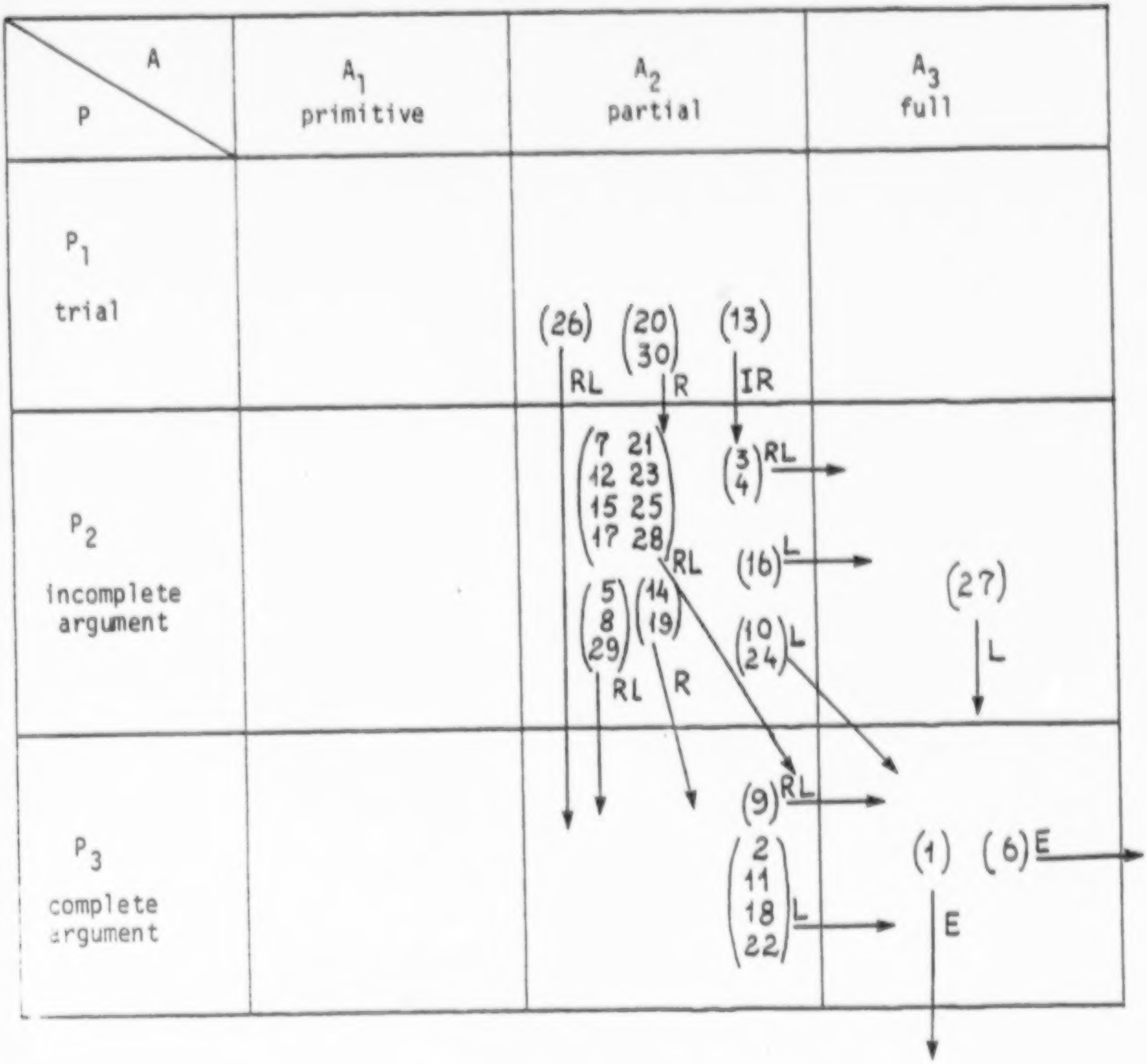

A
P
A_1 primitive
A_2 partial
A_3 full
P_1 trial
P_2 incomplete argument
P_3 complete argument
(26)
RL
(20 30)
R
(13)
IR
(7 21 12 23 15 25 17 28)
RL
(3 4)
RL
(16)
L
(27)
L
(5 8 29)
RL
(14 19)
R
(10 24)
L
(9)
RL
(2 11 18 22)
L
(1)
E
(6)
E

Figure 11 : Hints and progress for assignment 1

Many students gave an original A_2/P_2 response ("odd numbers") with justification ("three times an odd is odd; odd plus one is even"). To the L hint most responded "yes" and some added "I mean that there are an infinity of such numbers". Faced with the ineffectiveness of the L hint, we descended to the RL level, which helped almost every one in this cell. They easily proved that even numbers do not fit. A group of 8 students (grouped together in the table) also found some non-integer answers but did not generalize.

With regard to non-integers, it is worth comparing the A_2/P_2 cell with the two vertically adjacent cells. Students in A_2/P_1, when asked about non-integers, said that it is impossible to obtain an even number from a fraction. Students in A_2/P_2 were apologetic for not having checked fractions but, as a rule, could not generalize without further help. Whereas students in A_2/P_3 (except for one) generalized very quickly.

In conclusion, this table gives a first indication that the categorization scheme is more than technical, it reflects qualities of understanding and abilities to make effective use of different levels of hints.

Assignment 2: *Given the expressions $x^2 - y^2$ and $(x - y)^2$. Find pairs (x, y) which when substituted in the above expressions give the same result.*

The main categories on the Answer scale were:

A_1 : (at most) numerical examples for $x = y$ only, or $y = 0$ only;

A_2 : the generalization $(x, y)/x = y$ only, or $(x, y)/y = 0$ only, but no further;

A_3 : complete generalization $(x, y)/x = y$ or $y = 0$.

The main categories on the Procedure scale were:

P_1 : checking numerical examples or algebraic manipulation of expressions;

P_2 : proof of sufficiency;

P_3 : proof of sufficiency and necessity.

The following are examples of hints used:

I - What do you get if you substitute in both expressions $x = 1$, $y = 1$; $x = 1$, $y = 0$?

- Try $x = 5$, $y = ?$; $x = ?$, $y = 0$.

IR - If $x = 3$, can you find values for y, so that the results will be the same?

R - Are you sure that all the pairs in your answer satisfy the condition?

RL - Can you get equality for other values of x and y?

- For which values of the two expressions can we possibly get equality?

L - Are you sure that you gave a correct and complete response?

Figure 12 illustrates student progress as a result of the first effective hint.

One third of the students were classified in A_1/P_2 or A_2/P_1. Those in A_1/P_2 responded "all x, y, since the expressions are the same". To the L hint they answered "yes", to RL they responded "one can't tell", not until an R hint given was any progress made.

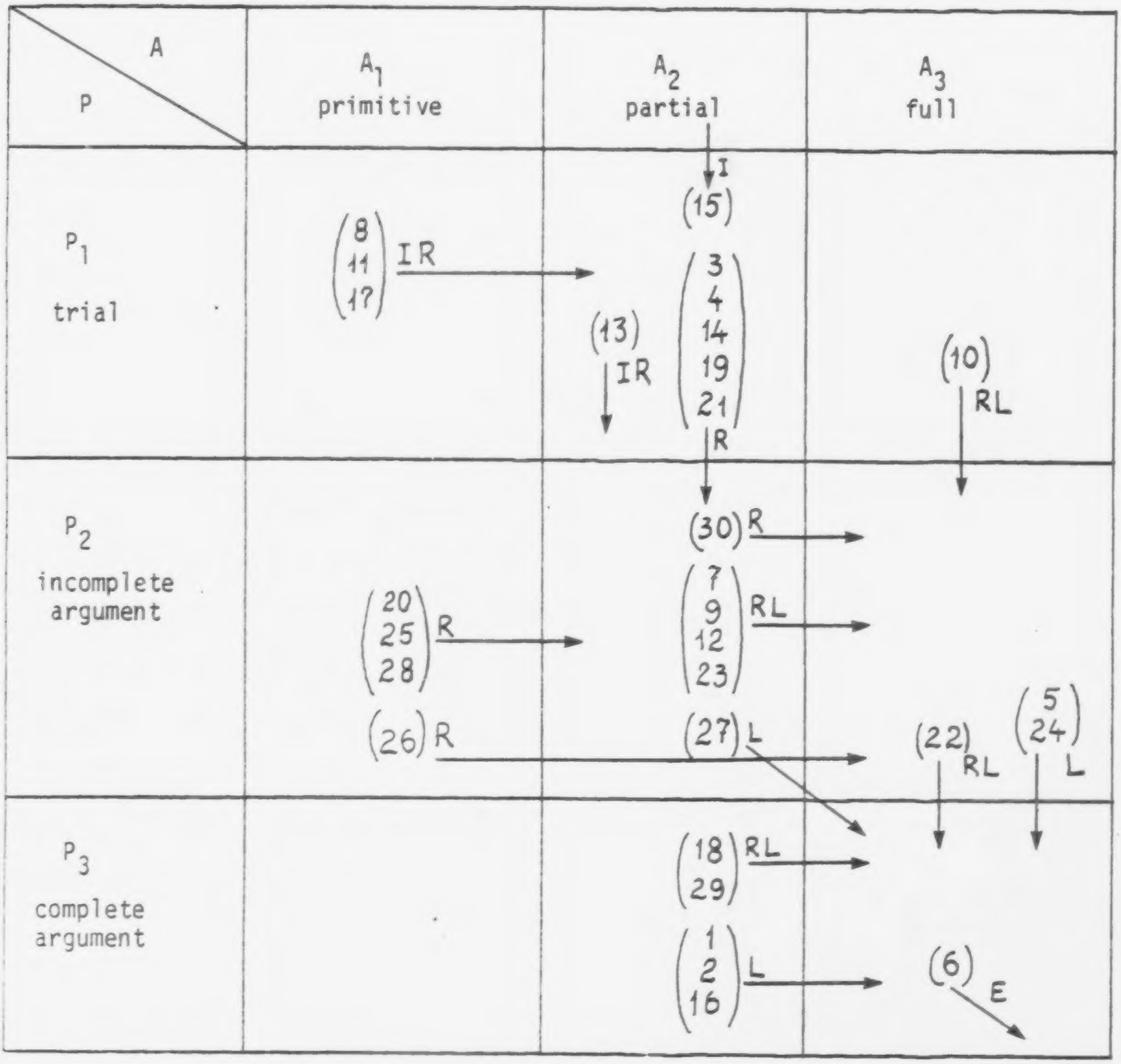

A
P
A1 primitive
A2 partial
A3 full
P1 trial
P2 incomplete argument
P3 complete argument
I
(15)
8 11 17
IR
3 4 14 19 21
(13)
IR
R
(10)
RL
(30)
R
7 9 12 23
RL
20 25 28
R
(26) R
(27) L
(22)
RL
5 24
L
18 29
RL
1 2 16
L
(6)
E

Figure 12 : Hints and progress for assignment 2

Students in A_2/P_1 gave one generalization (one gave $y = 0$ and the rest $x = y$). To the L hint they answered: "No, there are an infinity of pairs and I gave only a few examples", indicating that they had not understood their own generalization. This was reinforced by the R hint which caused them to check again a pair or two and then to "prove" their conclusion.

Again we see that the quality of effective hint is related to the quality of the original response (understanding).

Assignment 3: *Given the equation $b(2x + b) = 175$, where x and b represent positive integers; what can you say about x and b?*

A_1 : (at most) one value for b and one for x;

A_2 : two answers;

A_3 : the full answer:

$$b = 1 \quad x = 87, \quad b = 5 \quad x = 15, \quad b = 7 \quad x = 9.$$

The main categories on the Procedure scale were:

P_1 : numerical substitution and/or algebraic manipulation;

P_2 : some general conclusion; e.g., b divides 175;

P_3 : as in P_2 together with the use of the fact that b and x are positive integers.

The following are examples of hints used:

I : Substitute for b some of the numbers, 3, 4, 5, 6, 7 and find x.

IR : What can you deduce from the fact that 175 is equal to the product of two factors?

R : What is the relation between b and 175?

RL : Have you used all the data?

L : Are you sure that you have found all possible answers?

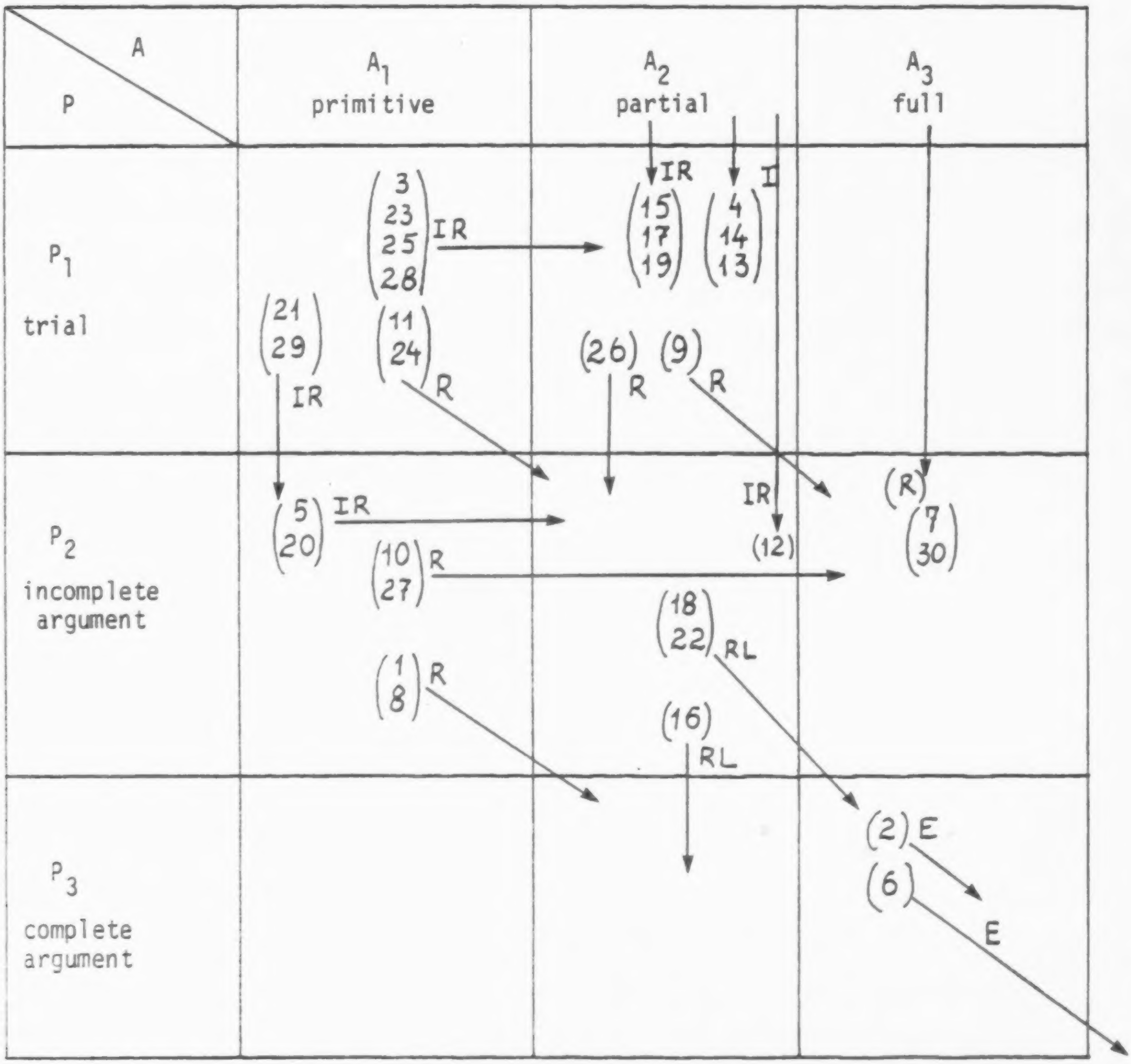

A
P
A1 primitive
A2 partial
A3 full
P1 trial
P2 incomplete argument
P3 complete argument
3 23 25 28
IR
15 17 19
4 14 13
I
21 29
IR
11 24
R
(26)
R
(9)
R
5 20
IR
10 27
R
(12)
(R)
7 30
18 22
RL
1 8
R
(16)
RL
(2) E
(6)
E

Figure 13: Hints and progress for assignment 3

The students who did not attempt this assignment said it was so unfamiliar that they did not know how to start. In general, these and the students in the A_1/P_1 primitive cell, were helped either by I or IR hints. The findings described above show a certain regularity. In general, students completely outside the table needed an instrumental hint to get started; those in A_1/P_1 an instrumental relational hint; those in A_2/P_1 or A_1/P_2 a relational hint and so on. If we use the expression $H > A/P$, where the level of hint is above that suggested by the above observed regularity with respect to the relevant A/P cell, and $H = A/P$, $H < A/P$ similarly, then the relevant percentages for all three items are given in Table 9.

	H < A/P	H = A/P	H > A/P
Assignment 1	13.3%	73.4%	13.3%
Assignment 2	16.7%	80.0%	3.3%
Assignment 3	10.0%	63.3%	26.7%
Total	13.4%	72.2%	14.4%

Table 9 : The relation between H and A/P as percentages of the responses.

If we look at the results of all three assignments, 72% exhibited this natural relation between the first useful hint and the quality of solution. The remainder were divided almost equally between those who responded to a hint at a higher level than normal and those who needed a lower level hint than might have been expected.

This is reasonable confirmation of the validity of the classification as well as of the effectiveness of the guidance system.

7.2.2. Diagnostic evaluation

In order to make progress on an assignment a student first has to *understand* that there is something left to be done, then he has to *want* to go on working, and last but not least, he needs to *know* how to proceed. Thus we can, in general, identify three types of reason for a student having stopped in a particular cell: does not *understand, understands* but does not *want, understands* and *wants* but does not *know* how.

The last case is where the empirical results described above, indicate that the classification/hint pattern exists. Those cases which do not fit this pattern give new insight to interesting difficulties. For example, in assignment 1 (Figure 11), S(26) did not give any explanation for the answer "all odd numbers". To the logical hint he just said "yes" but to the relational-logical hint (which is at a higher level than expected) he reacted nicely: "You probably want me to give a proof. I checked some numbers in my head and thought it would be better to prove". Not very willingly, he gave a sufficient and necessary proof for integers and stopped. The student *understood* that there was something left to be done, *knew* how to do it - but did not *want* to.

Another such example is S (10). Her solution was classified A_2/P_2, yet she responded to the logical hint: "If you meant integers my answer is complete. As for fractions, they confuse me. I did not feel like trying them". In the interview she moved to A_3/P_3. These two cases suggest that there may be affective aspects interacting with the cognitive. We shall refer to such difficulties as *cogno-affective*.

The situation where a hint at a lower level than expected was needed, is also of interest. According to the pattern, S (19) should have been able to profit from a relational-logical hint. Instead, in response to that type of hint, she said, "They asked me to substitute and not what not to substitute". This demonstrates the difficulty in *understanding* the need for proving a necessary condition (completeness) which is strictly cognitive.

In assignment 2 (Figure 12), S(27) reacted quickly to a logical hint and completed his response. Being asked why he stopped in the first place, he said: "I found a whole set of pairs, proved that my answer was correct, was satisfied and did not bother to look for more, unfortunately". This seems to us a cogno-affective reason, and fits the suggestion of the examples in assignment 1 that, in such cases, a higher hint may be sufficient.

S(30), on the other hand, was trapped by his proof. He assumed that for the expressions to be equal, their value must be zero and thus x must be equal to y. The relational-logical hint was of no use to him, but a relational one ("There are more values of x and y.") led to progress. Another case where a lower hint was needed is S(18). He originally wrote

$$\begin{aligned} (x - y)^2 &= x^2 - y^2 \\ x^2 - 2xy + y^2 &= x^2 - y^2 \\ -2xy &= 2y^2 \\ xy &= y^2 \\ x &= y \end{aligned}$$

To the logical hint he responded: "Well, I first checked examples and found that for $x = y$ both expressions are 0. So I assumed that this was the right answer and proved it." Given a relational-logical hint he discovered the $y = 0$ answer, but was in dilemma - why did he not get this answer from the equation? Here was a pure cognitive obstacle - the awareness that cancelling in algebra (as opposed to arithmetic) can lose information (see also in 6.3.)

The data for assignment 3 (Figure 13) further supports the suggestion that a student who benefits from a hint at a higher level than normal, is likely to have what we have called a cogno-affective difficulty. On the other hand, a student who needs a lower hint, is likely to have a specific cognitive difficulty. Thus, for example, S(7), who had avoided the problem altogether, because it seemed unfamiliar, responded to a relational hint, understood that b must be a factor of 175, and gave an almost complete response. On the other hand the relational hint did not help S(5), who had responded with $x = \frac{175 - b^2}{2b}$. He did not know what to do with the expression he had found, and concluded that one cannot find any specific b or x, only an expression for one in terms of the other, because there is only one equation in two unknowns.

He needed an instrumental-relational hint.

The main purpose of the "clinical" application was to achieve some general contribution to the diagnostic evaluation of 'high-level' mathematical activity, which will be applicable in the regular classroom setting. The discussion above suggests that the A/P table classification can be significant in the understanding of student achievement in problem-solving, and the limits thereof. The cells in the table provide us with a classification of qualities of mathematical activity and suggest the associated hierarchy of hints designed to reactivate the student from the point at which he stopped. The hierarchy of hints was tested in student interviews and its subsequent analysis established the connection with the A/P response classification. The interviews also indicated types in which the connection between hint and response can be profitably modified. Thus in cases of cogno-affective difficulties (e.g. maths anxiety, fractions confuse, "don't like proving") a higher level hint than usual can be useful and desirable. The reverse situation may occur when there are "classical" cognitive difficulties (e.g. necessary conditions, cancelling zero, one equation in two variables).

In the goal-framework, we defined qualities of mathematical activity and, at this stage, a guidance hierarchy connected to the goal framework has been established. A further step is to use the diagnostic information in practical implementation, which will be described in the following section.

7.3. Communication study (1982)

It was clear to us that teachers need more preparation before they are ready to use effectively the assignment projects in their classes. The premise was that we could communicate better with the teachers in terms of the qualitative evaluation and the guidance hierarchy, if we offered the research-instrument as a treatment tool, because it is shorter and more structured than the projects. A study was designed to examine whether the established method of evaluation and guidance works practically in class and whether it is effective. It was decided to use the two versions of the assignment tests that had been used in the main comparative study together with evaluation schemes (see 5.5.1.) and switch them between two groups. This design of the study was meant to check the assumption (cf. 5.5.1.) that the two versions are of similar difficulty for all three assignments and thus could be matched. This information could provide more validation to the finding in the main study and to the comparison planned for this study.

The plan of the study was that students would do an assignment worksheet and as soon as possible thereafter, would receive their papers with hints as suggested by the established model described in 7.2. The expectation was that they would improve their solutions using the hints provided. A month later they would do another assignment worksheet.

We were interested in the following pedagogic questions:

(a) Are the teachers able to operate the treatment without further help?

(b) Is it effective? Is there a "population shift" in the desired direction from the first to the second worksheet?

(c) What is the difference in terms of "population shift" between such a "short" treatment and the "long" treatment given in the main study?

We were also interested in the following technical questions:

(d) What is the extent of objectivity of the evaluation schemes used by the teachers?

(e) How much are the two versions of the assignment test the same, in "before" and "after" treatment situations?

Six teachers who expressed their willingness to participate and their ninth grade A stream classes were involved. They were divided into two matched ability groups (GR1, n=84 and GR2, n=83). This study was performed at the beginning of the school year when the material of the achievement test was not yet covered. Thus the matching was based on a measure of general reasoning factor - the *Necessary Arithmetic Operations* test (RG - 3, NAO) obtained from Ekstron et al. (1976). The average NAO scores for the six classes and the two groups are given in Table 10.

	n	NAO	
Class 1	31	60.2	GR1
Class 2	33	53.8	n = 84
Class 3	20	47.6	$\overline{\text{NAO}}$ = 54.7 σ = 17.4
Class 4	32	55.5	GR2
Class 5	25	53.7	n = 83
Class 6	26	53.1	$\overline{\text{NAO}}$ = 54.2 σ = 16.2

Table 10 : Average NAO scores for the two groups.

Both groups worked during one class period on an assignment test. GR1 was given the pretest version (PR) and GR2 the posttest version (PS). Since in this study the second application was intended to be after one month, the second assignment of the posttest version was not the same as in the pretest, as before, but was changed to the following:

Given the expressions $(x - 2)(y - 10)$ and $(2 - x)(y - 4)$, find pairs (x, y) which when substituted in the above expressions give the same result.

Student papers were checked independently by the class teacher and the researcher using the evaluation schemes. The teachers felt comfortable with the tables and some of them were very enthusiastic about the classification method and its objectivity. The percentages of agreement between teacher and researcher are given in Table 11.

	Agreement* Item	(a)	(b)	(c)	(d)
GR1	Assignment 1	84.7	7.1	3.5	4.7
	Assignment 2	84.5	8.3	3.6	3.6
	Assignment 3	76.3	5.9	8.3	9.5
GR2	Assignment 1	73.7	9.6	6.0	10.7
	Assignment 2	85.6	7.2	2.4	4.8
	Assignment 3	80.8	6.0	4.8	8.4

* (a) full agreement

(b) agree on main category and disagree on refined classification

(c) disagree on main category

(d) teacher cannot decide on classification.

Table 11 : Agreement percentages of the classification

The hierarchy of guidance obtained from the diagnostic study was carefully explained to the teachers. Hints to all solutions were written by the teacher together with the researcher as indicated by the model. Within one week of their doing the assignments, the students received their papers with the hints. A lesson in class was devoted to improving the solutions by the students using the suggested hints. The teachers provided further guidance when needed and students were advised to finish their work at home. A month later GR1 students took the (PS) version and GR2 did the (PR) version.

For all the students the "population shift" was measured by the same methods as used in the comparative study.

We could now answer the questions of the study (The findings are given in Appendix E in full).

(a) The teachers were able to use the evaluation scheme but had difficulties providing the hints, especially in cases when their previous acquaintance with the students caused them to suspect serious cognitive or cogno-affective difficulties of the sort we observed in the clinical study (section 7.2.)

(b) There was a definite improvement in the quality of student responses, as expressed in the A/P tables, from the first to the second application of the assignments.

(c) The improvement was not as great as in the main study. This is only to be expected because of the different form of the study; both the treatment and the period of time devoted to it were on a smaller scale.

The answers to the technical questions gave more support to the validity of the research instruments used in the main study.

(d) Full agreement between teachers and researcher was obtained for more than 80% of the cases. It seems that, after remodifying the schemes as result of this application, it will be impossible to get much higher agreement, even if evaluation will remain subjective to a certain extent.

(e) The study was designed so that two matched ability groups worked, one on the pretest version and the other on the posttest version, before the treatment. Generally the two versions were found to be of about the same difficulty.

The two versions were switched between the groups after the treatment and the improvement in both was also quite similar. The comparison between the two versions can be obtained from the tables in Appendix E. An alternative presentation of the data, which will be introduced in the following section (Table 14), will make the comparison even clearer.

In the implementation of the guidance method two main problems arose. The first, already mentioned above, is modification of the hint-hierarchy for exceptional difficulties suspected by the teachers. The second is concerned with the work in class; the teachers wanted to group students while working on the annotated papers, because it was difficult for them to provide all the guidance requested in class. Are there any effective criteria for grouping in such a situation?

7.4. Interpreting model leading to further studies

In order to try to give answers to the questions which arose, the diagnostic findings were now used to interpret the data from the main comparative study and the communication study. The hierarchy of guidance which was established is summarized in Figure 14,where in each cell the hint that is likely to cause progress appears excluding the "special" cases for which $H < A/P$ or $H > A/P$.

Instrumental	A_1 primitive	A_2 partial	A_3 full
P_1 trial	*Instrumental-Relational*	*Relational*	*Relational-Logical*
P_2 incomplete argument	*Relational*	*Relational-Logical*	*Logical*
P_3 complete argument		*Logical*	*Extension*

Figure 14 : Answer - Procedure - Hint relations

In the data of the 1980 study, students' Q scores were calculated as the sum of $i + j$ for a cell A_i/P_j (see page 35). Based on the diagnostic findings as summarized in Figure 14, we establish a 3 - category link between the quality of solution and the guidance. Thus we categorize responses for which $Q < 2$ as *instrumental* responses-I, for $Q = 3, 4$ as *relational* responses- R and for $Q = 5, 6$ as *logical* responses- L .

The Q score for the test gives only some general information about student performance. A better evaluation of student work can be obtained if we consider all three assignments in the following way. We can represent each student paper by an ordered triple of three letters (using I, R, L), corresponding to his responses on the three assignments respectively. Such a triple gives a *profile* of student performance on the test. Since the assignments go from easy to hard we would expect the letters of the triple to be in decreasing or more precisely in non-increasing order. Each student was assigned a profile and the frequencies of the various profiles were calculated. At the pretest stage five different profile clusters were clearly distinguishable, after the rare ones (which occured for less than 5%) were omitted. At the posttest stage five "higher" profiles appeared but only for the experimental group. Table 12 exhibits the frequencies of the ordered profiles for the experimental and control groups in both applications of the assignment tests.

Group / Profile	Experimental group n = 191		Control group n = 194	
	pretest	posttest	pretest	posttest
RII	21%	10%	18%	21%
RRI	27%	13%	36%	37%
RIR	6%		5%	5%
RRR	21%	17%	18%	13%
LRI	5%	5%	6%	5%
RLI		10%		
LRR		5%		
RLR		9%		
LLR		7%		
LLL		10%		
	80%	86%	83%	81%

Table 12 : Profile frequencies for the experimental and control groups

The similarity of the profile frequencies for the two groups at the pretest stage, would seem to give further validation to the main instrument of the study and/or to the matching of the two groups. As would be expected, the control group looks much the same on the two versions of the assignment test. The improvement in the experimental group is seen very clearly in terms of the profiles.

We can consider a profile to be "regular" if there is some decrease in the order of the letters (except for LLL, which must also be regarded as "regular"). Thus all the profiles in Table 12 are regular, except for four: RIR, RRR, RLI and RLR. It is interesting to find out whether such profiles are peculiar also in terms of other parameters of the study. In the first place, average scores were calculated for the whole population at the pretest stage. The findings are given in Table 13.

Test / Profile	ACH		ATSI		ATMA	
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
RII 20%	51.4	(19.4)	60.8	(12.8)	61.9	(11.8)
RRI 31%	54.6	(18.1)	66.8	(11.2)	67.2	(10.4)
RIR 6%	48.6	(18.6)	59.6	(15.7)	68.7	(13.9)
RRR 20%	66.2	(18.5)	74.4	(14.6)	69.9	(12.4)
LRI 5%	68.8	(16.1)	73.4	(11.1)	72.5	(10.5)

82%

N=385

Table 13 : Average scores for profiles on the whole population (pretest)

Most of the figures in the table are what one might expect, since the Q score was found to be in high correlation with all other variables (see Table 6). Yet the RIR and RRR profiles show some peculiarity. The first one occurred for quite a small group, which had a low average on the achievement test relative to their place in the profile ranking.

Also, for the three regular profiles, the average scores on the ATSI and ATMA are almost the same, whereas for the RIR and RRR profiles there seems to be a difference: For RIR, the ATMA average score is larger than the ATSI and for RRR, ATSI score is somewhat larger than ATMA.

At the posttest stage the average scores were calculated for the profiles of each group separately. The results for the control group were very similar to those in the pretest. For the experimental group the RIR profile occurred for less than 5% and another two small "irregular" profiles were found, RLI and RLR, for which the ATSI was about the same as the AMTA average score. For the RRR profile once again the ATMA average score was somewhat greater than the ATMA (70.6, 68.1). A further clinical investigation of students having "irregular" profiles may lead to the modification of the guidance system in cases of specific cognitive and cogno-affective difficulties.

The profile clusters pattern was repeated for the data from the communication study. The profile frequencies are shown in Table 14.

Group Profile	GR1 n=84		GR2 n=83	
	(PR) version	(PS) version	(PS) version	(PR) version
RII	17%	12%	25%	10%
RRI	37%	24%	31%	18%
RIR	5%		7%	5%
RRR	14%	15%	14%	14%
LRI	7%	5%	8%	8%
RLI		12%		11%
LRR				7%
RLR		6%		5%
LLR		5%		
LLL		6%	5%	7%
	80%	87%	88%	85%

Table 14 : Profile frequencies for the two groups on both applications

The repetition of the findings presented in Tables 12 and 14 would suggest that we could try to group students to work on assignments according to these profiles. We have seen that there is relation between the order of the profile in the list and the average achievement level, but within each profile there were students with a variety of abilities. An empirical study to examine the effectiveness of profile grouping, compared to other methods, such as mixed groups or homogeneous ability groups, can help in planning the implementation of class work on assignments.

The most interesting aspect of our studies is after all the effectiveness of the treatment. A clear picture is obtained by graphical illustration of the findings for the (PS) version on both the main study (1980) and the communication study (1982).

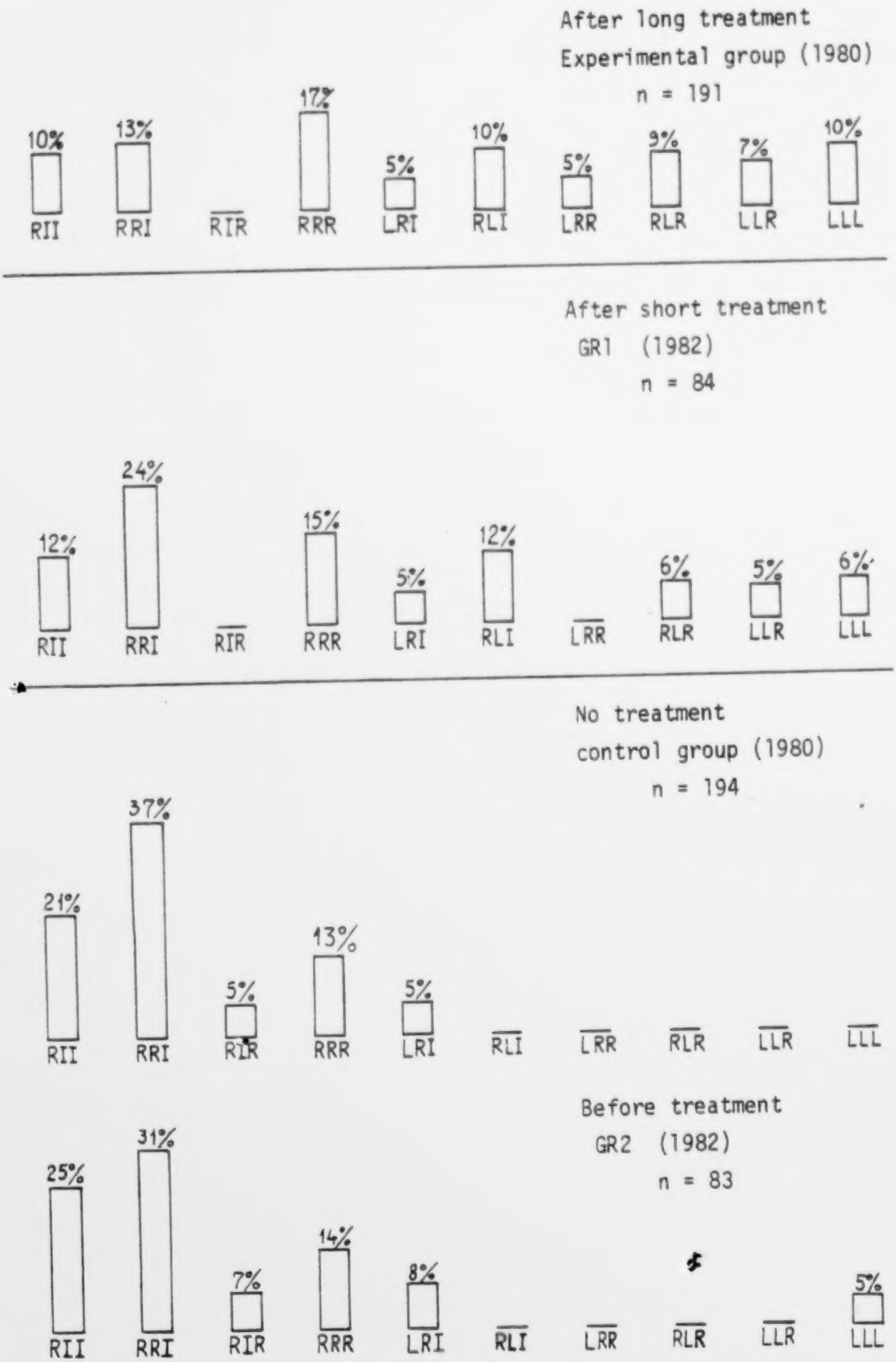
First, we can see from Figure 15, the similarity between the two groups without the treatment. Then we see improvement after the short treatment using the assignment test as a teaching instrument. Finally, we can see the greater improvement after doing the assignment projects over a longer period. Questions for further studies are:

(1) How effective is a series of short treatments?

(2) Will students benefit from teacher accumulated experience with the treatment?

(3) Comparison of different types of assignment worksheets; e.g. three separate items versus a three-part assignment.

(4) How will class experience with the guidance method on assignment worksheets affect work on the assignment projects themselves?



After long treatment
Experimental group (1980)
n = 191

After short treatment
GR1 (1982)
n = 84

No treatment
control group (1980)
n = 194

Before treatment
GR2 (1982)
n = 83

Figure 15 : Profile findings related to treatment

8. CONCLUDING REMARKS

This thesis is a summary of a study that has taken more than six years. The length of the study is undoubtedly due, in part, to the fact that we have tried to explore the relatively uncharted area of "high-level" mathematical activity, and have tried to establish both effective materials for and an effective practical diagnostic framework to encourage and enhance such activity. Thus the work was designed to explore and enhance student mathematical activity in open problem solving situations. In the first place we developed assignment projects for 9th grade students. We use the term *assignment* for a mathematical problem which has, to some degree, each of the following elements, and the potential to fulfill some of the following expectations:

(a) Although based on the curriculum already familiar to the student, it has some aspect of unfamiliarity. The expectation is that the student will be encouraged by the context of the problem to overcome the unfamiliarity.

(b) Various levels of possible answers and procedures. A student may obtain easily a primitive (partial) answer, but we expect that he will make a serious attempt to achieve a full answer.

(c) "Non-explicit invitation" to extend the answer, the procedure, and even the problem itself.

It seems to us that, by posing such assignments, we not only encourage the students to explore, but can learn a great deal about their work at the high ability levels.

It is significant to note that the content of the particular problems developed is integral to the main core curriculum and not, as so often in the literature, problems detached from any context and/or not linked directly to a core curriculum. This self-imposed restriction made the development of the projects more difficult, but we considered it to be necessary for a number of reasons.

If we want the improvement of student mathematical activity to be directly meaningful to the students' everyday mathematical activity, it is essential that the context of the questions should also be related to the regular curriculum.

It is clear that the goals for the students should be aimed at the teachers as well, if we want success with students. Open ended problems related to the core material are more likely to be used by the teachers. It is hoped that the experience teachers get when working on such projects in their classes will naturally affect their style of teaching mathematics and, consequently, enhance student mathematical activity.

The simple provision of problem material would not seem to be in itself a sufficient basis for the desired development. In order to guide the students to progress gradually, not only "good intuition" is desirable, but also some organised framework. This is even more desirable if we want to communicate to teachers how such activity might be organised and encouraged.

Studies were carried out to investigate whether in fact the material elicits the desired goals. This was followed by an evaluation of its effectiveness in enhancing student mathematical activity.

A major problem was the provision of relevant research instruments, since the study by its very nature focuses on "macro" rather than "micro" activities. Assignment tests were developed and validated in several ways.

Comparative studies were conducted in schools. The comparative evaluation was done under experimental conditions with actual classes, but, in addition to the regular teacher, there was strong intervention by the experimenters. The results of the comparative studies have shown that the assignment projects can indeed be effective in enhancing student mathematical activity. However, our experience indicated that if we want the material to be used effectively by the teachers, it is very important to provide a practical method of evaluating student work on such "open problems" and hence the means by which they can help students with suitable guidance. Also, it was clear that the treatment was too "heavy" and demanding for teachers and students. An appropriate preparation on a smaller scale might help.

Therefore, two follow-up studies were undertaken. The first one, a clinical study in which one class was chosen and all the students interviewed about their work on assignments within the goal-framework. This led to the establishment of a diagnostic classification of qualities of mathematical activity and suggested an associated hierarchy of guidance. The applicability of the diagnostic guidance and the effectiveness of a short treatment were examined successfully in the second (communication) study.

The qualitative evaluation model was further developed with the interpretation of the data from the comparative and communication studies in the light of the diagnostic findings. The result was the formation of clear profile clusters in terms of instrumental/relational/logical activity.

No doctorate thesis in mathematics education can ever be considered to have exhausted its subject, especially one devoted to the extensive field of higher-level activities. Some of the major questions and developments which are directly suggested by this work would seem to be:

(a) A clinical investigation of students within profile clusters, in order to obtain modification of the classification and guidance hierarchy in relation to cognitive and cogno-affective goals.

(b) An implementation study of optimal arrangement of class work on assignments considering the profiles.

(c) A study of the effect of a series of short treatments. Within the regular classroom and curriculum, a teacher can at intervals give his students assignments, classify their solutions in Answer/Procedure tables, and then to each response, he can write a suitable hint using the Instrumental to Extension hierarchy. Subsequently, in the class, he is free to guide students work and provide further hints as necessary and appropriate. If this process is repeated how will the quality of student response improve? How will the instruction toward higher-level activities improve?

We opened this thesis with Hawkins' (1973) words. We quote him again in conclusion:

> *The teacher himself is potentially the best researcher, if only we would offer him strong intellectual support and respect his potentialities as a scientist.*

In order to help the student to become "problem-researcher" - it would seem to help if the teacher himself became a researcher of problem solving.

BIBLIOGRAPHY

Bell, A.W. The learning of process aspects of mathematics. Educational Studies in Mathematics 1979, 10, 361-387.

Bell, A.W. and Shiu, C. Evaluating process aspects of a mathematical curriculum. Proceedings PME Fifth Conference, Grenoble, France, 1981.

Beth, E.W. and Piaget, J. Mathematical Epistemology and Psychology. Holland: Reidel, 1966.

Biggs, J.B. and Collis, K.F. Evaluating the Quality of Learning: The SOLO Taxonomy. New York: Academic Press, 1982.

Bloom, B.C. and Engelhart, M.D. (eds.). Taxonomy of Educational Objectives: The Classification of Educational Goals. I: Cognitive Domain. New York: D. McKay Co., 1956.

Bloom, B.C., Hasting, J.T. and Madaus, G.F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Bruckheimer, M. and Hershkowitz, R. Mathematics projects in junior high school. Mathematics Teacher, 1977, 70(7), 573-584.

Burton, L. The teaching of mathematics to young children using a problem solving approach. Educational Studies in Mathematics, 1980, 11, 43-58.

Cobb, P. and Steffe, L.P. The constructive researcher as teacher and model builder. Journal for Research in Mathematics Education, 1983, 14, 83-94.

Cronbach, L.J. Essentials of Psychological Testing (2nd edition) New York: Harper and Row, 1960.

Cureton, E.E. The normal approximation to the signed-rank sampling distribution when zero differences are present. J. Am. Stat. Ass. 1967, 62, 1068-1069.

Ekstron, R.B., French, J.W., Harman, H.H. and Derman, D. Manual for kit of factor-referenced cognitive tests. Princeton: Educational Testing Service, 1976.

Hadamard, J. An essay on "The Psychology of Invention in the Mathematics Field". New York: Dover, 1954.

Halmos, P. The heart of Mathematics. American Mathematical Monthly, 1980, 87, 519-524.

Hart, K. Children's understanding of mathematics: 11-16. London: Murray, 1981.

Hart, K. Tell me what you are doing - discussions with teachers and children. Proceedings PME Seventh Conference, Israel, 1983.

Hawkins, D. Nature, man and mathematics. In A.G. Howson (Ed.), Developments in mathematical education. Cambridge University Press, 1973.

Kilpatrick, J. Research on problem solving in mathematics. School Science and Mathematics, 1978, LXXVIII (3), 189-192.

Kilpatrick, J. Methods and results of evaluation with respect to mathematics education, in New Trends in Mathematics Teaching, Vol. IV, UNESCO, 1979.

Krutetskii, V.A. The Psychology of Mathematical Abilities in School Children. Chicago: University of Chicago Press, 1976.

Landa, L.N. Algorithmization in Learning and Instruction. Englewood Cliffs, NJ: Educational Technology Publication, 1974.

Lienert, G.A. Verteilungsfreie Methoden in der Biostatistik (Band II). Verlag Anton Hain. Meisensheim am Glan, 1978.

Mathematics Group. Assignment projects for ninth grade (booklets for A, B, C levels and teacher guides). Rehovot, Israel: Science Teaching Department, Weizmann Institute, 1975-1977.

NCTM (National Council of Teachers of Mathematics), An Agenda for Action: Recommendations for school mathematics of the 1980's Reston, Va: NCTM, 1980.

Newel, A., Simon, H.A. Human Problem Solving, NJ: Prentice-Hall, 1972.

Norman, D.A. (ed). Perspectives on Cognitive Science, Norwood, NJ: Ablex Publishing Corporation, 1981.

Polya, G. How to solve it. Princeton: University Press, 1945.

Polya, G. Mathematics and Plausible Reasoning, 2 Vols. Princeton: University Press, 1959.

Polya, G. Mathematical Discovery. New York: John Wiley, 1962.

Raatz, U. Die anwendung des Wilcoxon-tests fur paardifferenzen bei gruppierten Messwerten und grossem stichprobenumfang. Biometrical Journal, 1977, 19(3), 201-206.

Romberg, T.A., Collis, K.F., Donovan, B.F., Buchanan, A.E. & Romberg, M.N. A report on the NIE/ECS Item Development Project: The development of mathematical problem-solving superitems. Madison: Wisconsin Center for Education Research, 1982a.

Romberg, T.A., Jurdak, M.E., Collis, K.F., & Buchanan, A.E. A report on the NIE/ECS Item Development Project: Construct validity of a set of mathematical superitems. Madison: Wisconsin Center for Education Research, 1982b.

Scandura, J.M. Problem Solvers. London: Academic Press, 1977.

Schoen, H.L. and Oehmke, T.A. New approach to the measurement of problem solving skills, in Problem Solving in School Mathematics NCTM, 1980 yearbook.

Schoenfeld, A.H. Measures of problem solving performance and of problem solving instruction. Journal of Research in Mathematics Education, 1982, 13, 31-49.

Schoenfeld, A.H. Theoretical and pragmatic issues in the design of mathematical "problem solving" instruction. A paper presented at the 1983 Annual Meeting of the American Educational Research Association, Montreal, Canada, 1983.

Siegel, S. Nonparametric Statistics New York: McGraw-Hill, 1956.

Skemp, R.R. Intelligence, Learning and Action. Chichester: John Wiley, 1979a.

Skemp, R.R. Goals of Learning and Qualities of understanding. Mathematics Teaching, 1979b, 88, 44-49.

Wertheimer, M. Productive Thinking. San Francisco: Harper & Row, 1959.

Wheeler, D. Mathematization - Notes for a Study Group. Montreal: Concordia University (unpublished papers), 1978.

Zehavi, N. A final project for 9th grade: development, implementation and analysis. Proceedings PME Third Conference, Warwick, 1979

Zehavi, N. and Bruckheimer, M. Assignment test: an instrument for qualitative evaluation of student mathematical activity. Proceeding PME Sixth Conference, Antwerp, 1982a.

Zehavi, N. and Bruckheimer, M. Problem solving: linear functions and magic circles. International Journal for Mathematics Education in Science and Technology, 1982b, 13, 5, 619-625.

Zehavi, N. and Bruckheimer, M. $x^2 - y^2$ is equal to $(x - y)^2$. Mathematics in School, 1984, 13(1), 22.23.

Zehavi, N., Bruckheimer, M. and Ben-Zvi, R. Qualitative evaluation of mathematical activity and its relation to effective guidance. Educational Studies in Mathematics, in press.

APPENDICES

APPENDIX A

TEXT OF A SAMPLE OF ASSIGNMENT PROJECTS

Assignment Project: Squares (level I)

I. (a) Draw a square with side of length 3 cm. Extend each side by 2 cm to obtain a new square. Colour and compute the additional area.

(b) Draw a square and denote the length of a side by x cm (x represents a positive integer). Extend each side by 2 cm to obtain a new square.

(c) (i) Write down an open phrase for the area of the square with side x cm.

(ii) Write down an open phrase for the area of the new enlarged square.

(iii) Write down an open phrase for the difference in area between the two squares (i.e. for the additional area) and reduce to simplest form.

(d) Compute the additional areas for five different integer values of x.

(e) Which of the following statements are true and which are false? (Justify your answers.)

(i) The additional area is always represented by an odd number.

(ii) The additional area is always represented by an even number.

(iii) The additional area is always represented by a multiple of 4.

(iv) The additional area is always represented by a multiple of 8.

(f) If x is odd the additional area is a multiple of ___________.

If x is even the additional area is a multiple of ___________.

(g) Find the first four integer values of x for which the additional area is divisible by 16 and try to establish a pattern.

(h) Given that the additional area is 128 cm^2, what can you say about the side of the original square?

(i) If the side of the original square is increased, what can you say about the additional area?

(j) What is the smallest possible value of the additional area?

(k) Given that the additional area is more than 114 cm^2, what can you say about the side of the original square?

(ℓ) Given that the side of the original square is not greater than 40, what is the largest possible value of the additional area?

(m) Try to pose a further problem, similar to those in the previous sections and answer your problem.

II. We are given a square whose side, measured in cm, is an integer and denoted by x. Each side is extended by the same amount to form a new square.

(a) The open phrase for the additional area is $5(2x + 5)$. By how much was each side extended? Explain your answer.

(b) Draw a suitable figure and colour the additional area.

(c) Compute the additional area for three different integer values of x. Is there a common factor to all the results.

(d) Can you find integer values for x, so that the additional area is a multiple of 10.

(e) Write problems similar to those in sections (h) - (m) in Part I which apply to this case. Answer your problems.

III. Given a square whose length in cm is an integer denoted by x. By adding b cm to each side (b is a positive integer) a new square is obtained.

(a) Write down an open phrase for the additional area and show that it is a multiple of b.

(b) For which values of x is the additional area an odd/even multiple of b?

(c) Given that the additional area is 175 cm^2, what can you say about the side length of the original square?

(d) Given that the additional area is 432 cm^2, what can you say about the side length of the original square?

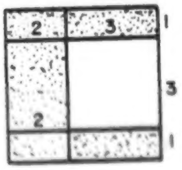
(e) (i) Write down an open phrase for the additional area of a square with side x cm, which is enlarged by 8 cm.

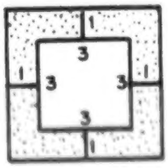
(ii) Write down an open phrase for the additional area of a square with side y cm, which is enlarged by 2 cm.

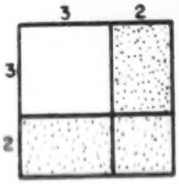
(iii) Write an open sentence which says that the additional areas in (i) and (ii) are equal and find its solutions.

Solutions to selected sections

I. (a) There are several possibilities for the extended square:

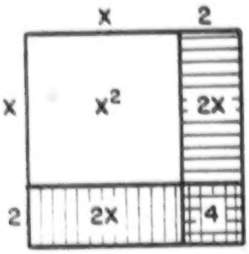


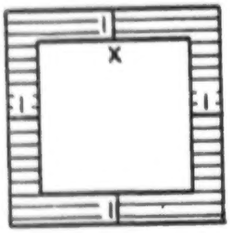




The additional area is 16 cm^2 : $5^2 - 3^2 = 16$.

(b) Some possible ways of drawing the extended square are:





(c) An open phrase for the additional area is : $(x + 2)^2 - x^2$.

It can be simplified in several ways:

(i) $(x + 2)^2 - x^2 = x^2 + 4x + 4 - x^2 = 4x + 4 = 4(x + 1)$

using $(a + b)^2 = a^2 + 2ab + b^2$.

(ii) $(x + 2)^2 - x^2 = (x + 2 - x)(x + 2 + x) = 2(2x + 2) = 4(x + 1)$

using $a^2 - b^2 = (a - b)(a + b)$.

(f) If x is odd the additional area is a multiple of 2, 4, 8.

If x is even the additional area is a multiple of 2, 4.

(g) The first four integer values for which the additional area is divisible by 16 are: 3, 7, 11, 15. The pattern is every second odd number starting from 3, or in a formal presentation:

$$4x + 4 = 16k$$

$$x = 4k - 1 .$$

(j) The smallest possible value for the additional area is 8 cm^2 and is obtained for $x = 1$, since the value of the expression $4(x + 1)$ increases with x.

(k)
$$\begin{aligned} 4x + 4 &> 114 \\ 4x &> 110 \\ x &> 27.5 \end{aligned}$$

The side of the original square is at least 28 cm.

II. In this part we start with an open phrase for the additional area and work back to the situation from which it arose.

(a) We bring here several possible responses at various cognitive levels, from the most primitive induction to a full deductive reasoning.

(1) A guess that the extension is 5 cm and then a check that it really gives $5(2x + 5)$.

(2) The extension is denoted by b and an open phrase $b(2x + b)$ is obtained.

One "can see" that b needs to be 5 in the given case.

(3) $2b + b^2 = 10x + 25$

hence $b^2 = 25$, i.e. $b = 5$

This is followed by a check that for $b = 5$ $\quad 2bx = 10x$.

(4) A backwards argument gives that

$$x^2 + 5(2x + 5) = (x + 5)^2,$$

hence the extension is 5 cm.

(5) $2bx + b^2 = 10x + 25$ must be true for any x.

Let's subtitute $x = 3$

$$b^2 + 6b - 55 = 0 .$$

The positive solution is 5 and the negative one does not fit the problem.

(6) A general solution:

$$b^2 + 2bx - 10x - 25 = 0$$

$$b_{1,2} = \frac{-2x \pm \sqrt{4(x + 5)^2}}{2}$$

The negative root is not suitable because, even if b represents a decrease in side length, it exceeds x.

Therefore, only the positive solution is suitable:

$$b = \frac{-2x + 2(x + 5)}{2} = 5 \text{ .}$$

III. (c) The related equation is $b(2x + b) = 175$ where x denotes the side length and b denotes the added length.

for $b = 1$ $\quad 2x + 1 = 175$ $\quad$ *thus* $\quad x = 87$

for $b = 5$ $\quad 2x + 5 = 35$ $\quad$ *thus* $\quad x = 15$

for $b = 7$ $\quad 2x + 7 = 25$ $\quad$ *thus* $\quad x = 9$

(d) Factorizing 432 gives $432 = 2^4 \cdot 3^3$

If b is odd then $2x + b$ is also odd and their product cannot be even (432). Thus b must be even and, hence, $2x + b$ is also even. The number 432 is an odd multiple of 2^4. The factor 2 divides $2x + b$ thus it appears in b no more than 3 times. But since $b < 2x + b$, b is bounded by $21(\sqrt{432} \cong 20.8)$ and the full solution is:

for $b = 2$ $\quad 2x + 2 = 216$ $\quad$ *thus* $\quad x = 107$

for $b = 4$ $\quad 2x + 4 = 108$ $\quad$ *thus* $\quad x = 52$

for $b = 8$ $\quad 2x + 8 = 54$ $\quad$ *thus* $\quad x = 23$

for $b = 6$ $\quad 2x + 6 = 72$ $\quad$ *thus* $\quad x = 33$

for $b = 12$ $2x + 12 = 36$ *thus* $x = 12$

for $b = 18$ $2x + 18 = 24$ *thus* $x = 3$.

(e) (i) $8(2x + 8)$

(ii) $2(2y + 2)$

(iii) $2(2y + 2) = 8(2x + 8)$

The equation can be simplified.

$$4y + 4 = 16x + 64$$

$$4y - 16x = 60$$

$$y - 4x = 15 .$$

All the solutions will lie on the graph of the line $y = 4x + 15$

A partial table is :

x	1	2	3	4	5	6	...
y	19	23	27	31	35	39	...
additional area	80	96	112	128	144	160	...

Note the regularity of the three series.

Assignment Project: Squares (level II)*

I. Given a square whose length in cm is an integer denoted by x. A new square is obtained by adding 5 cm to each of its sides.

(a) Draw a suitable diagram. Shade the additional area (the difference between the areas of the two squares).

(b) Write down an open phrase for the additional area and reduce it to its simplest form. Specify the domain (set of substitutions) relevant to the problem.

(c) Show that the additional area (in cm^2) is an odd multiple of 5.

(d) If you know that the additional area is more than $256cm^2$, what can you say about the side of the original square?

(e) What is the smallest possible value of the additional area?

(f) Is it possible to have a situation in which the additional area is an integer multiple of the area of the original square?

(g) (i) If x is 10, what is the percentage increase in area?

(ii) Same question, but x is 50 ?

(h) Find the side of the square for which the increase in area is closest to 100%?

(i) What can you say about the function which relates the side length of the original square to

(i) the additional area, (ii) the area of the original square,

(iii) the percentage increase in area.

(j) If instead of x an integer, we allow x to be any positive real number, discuss sections (b) - (i) accordingly.

---

* Acknowledgement to Rina Hershkowitz for her major contribution to this version.

II. What will happen if the side of the square decreases by 5 cm instead of increasing?

(a) Draw a suitable diagram and write down an open phrase for the difference in area.

(b) Write down a function which relates the side length to the difference between the areas of the two squares and draw its graph.

III. Now let's take a given square with side of length 4 cm and let the extension be a variable denoted by b.

(a) Write down a function which relates b to the difference in area and draw its graph.

(b) Write down a function which relates b to the absolute value of the difference in area.

IV. Let's modify our problem once again and look what happens if both the side x cm of the given square and the extension b cm are variables, with the condition that b is proportional to x. Explore the problem.

V. (a) Let's explore a new problem: Given a square of side x cm. The length of one pair of opposite sides is increased by b cm and that of the other pair decreased by b cm, to obtain a rectangle.

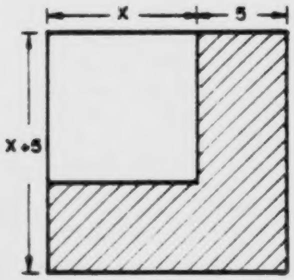
(b) The same as (a) with the condition that b is proportional to x.

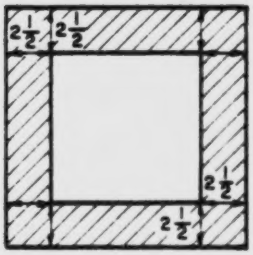
VI. Try to invent more problems dealing with the relation between changes in length and area.

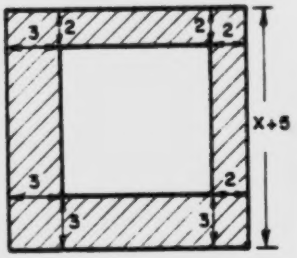
Solutions to selected sections

I. (a) There are several possibilities for drawing the extended square.

For example,






3
2
2
2
x+5
3
2
3
3

(b) (1) Algebraic approach

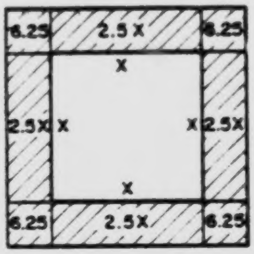
$$(x + 5)^2 - x^2 = x^2 + 10x + 25 - x^2$$

$$= 10x + 25$$

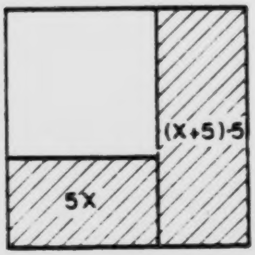
$$= 5(2x + 5).$$

(2) Geometric approach

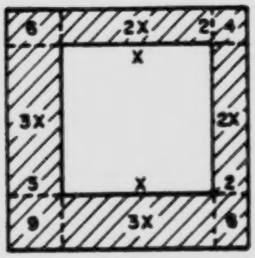
For example,


6.25
2.5x
6.25
x
2.5x
x
x
2.5x
x
6.25
2.5x
6.25

$$4 \cdot 2.5x + 4 \cdot 6.25 = 10x + 25$$
$$= 5(2x + 5)$$



$$5x + (x + 5) \cdot 5 = 10x + 25$$
$$= 5(2x + 5)$$


6
2x
2
4
x
3x
2x
3
x
2
9
3x
6

$$2 \cdot 2x + 2 \cdot 3x + 4 + 2 \cdot 6 + 9 = 10x + 25 = 5(2x + 5)$$

(d) $$10x + 25 > 256$$
$$10x > 231$$
$$x > 23.1$$

The side of the original square is at least 24 cm.

(e) The value of the expression $10x + 25$ increases as x increases and thus, the least value will be obtained for $x = 1$ and is 35 cm^2.

(f) The only two possibilities, $x = 1$ and $x = 5$, can be obtained by guess and check procedures.

We can also approach the problem systematically. For example:

$$k = \frac{10x + 25}{x^2}$$

(i) Let's try to guess possible values for x.

How many values are we going to need to check? Are we going to live long enough?

Since $10x + 25$ is an odd number, x^2 and therefore x must be also odd.

What happens to k as x increases? (k decreases)

When is $k < 1$? ($x > 13$).

So all we need to check is $1 < x < 11$, x odd.

(ii) In the above we investigated possible values for k. Alternatively we can investigate x.

$$kx^2 - 10x - 25 = 0$$

$$x = \frac{10 \pm \sqrt{100 + 100k}}{2k}$$

The positive root only is relevant

$$x = \frac{5}{k}(1 + \sqrt{1 + k}).$$

Possible values of k for which $1 + k$ is a perfect square are:

3, 8, 15, 24, 35 (for larger numbers $x < 1$)

Only 3 and 35 give an integer.

(iii) We can also find the answer by factorising the open sentence

$$kx^2 = 10x + 25$$

$$x(kx - 10) = 5 \cdot 5$$

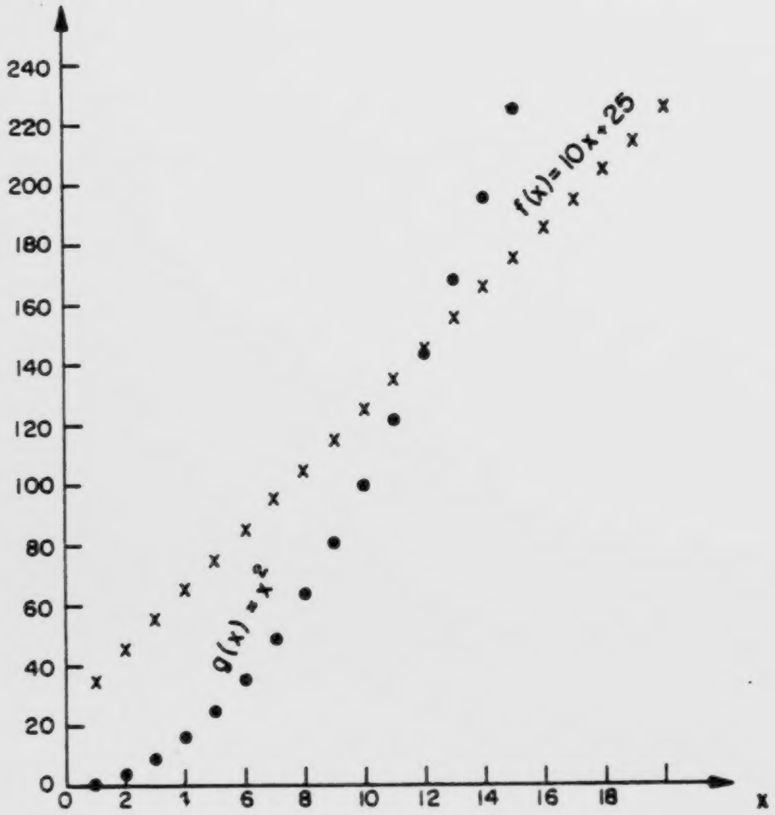
Thus x can be only 1 or 5 (or 25 which fails).

(g) For x = 10 the percentage increase is 125%

For x = 50 the percentage increase is 21%.

This section was meant to surprise. In the previous sections the students notice that the additional area increases with increasing x. Here the percentage *decreases* with increasing x.

(i) (i) $f(x) = 10x + 25$, (ii) $g(x) = x^2$

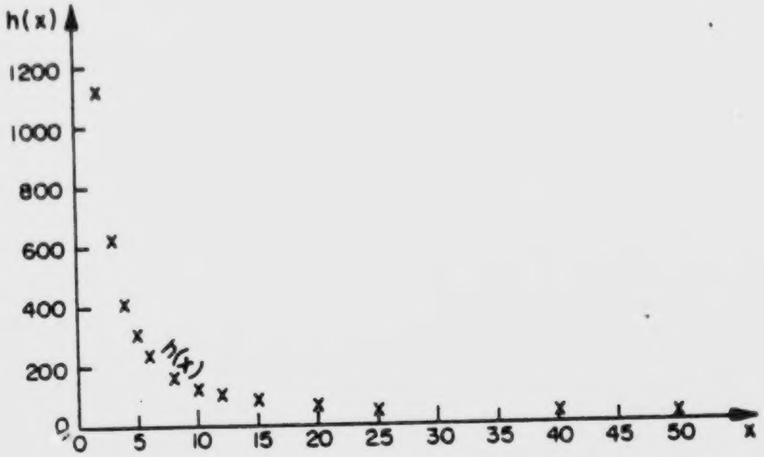


The function $f(x)$ is a discrete increasing function, with constant rate of increase of 10 cm^2. The function $g(x)$ is also a discrete increasing function, but the increase depends on the value of x

$$\frac{\Delta g}{\Delta x} = \frac{(x + 1)^2 - x^2}{x + 1 - x} = 2x + 1.$$

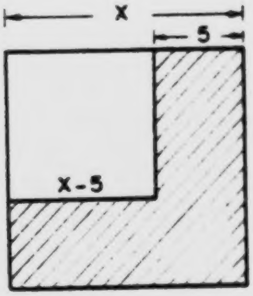
(iii)

$$h(x) = \frac{10x + 25}{x^2} \cdot 100 = \frac{f(x)}{g(x)} \cdot 100$$



The function $h(x)$ decreases.

II. (a) One possible drawing is::



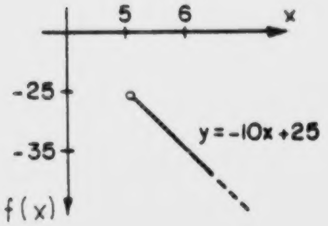
An open phrase for the difference in area is

$$(x - 5)^2 - x^2 = x^2 - 10x + 25 - x^2 = -10x + 25 .$$

The geometry of the problem implies that x must be greater than 5 and thus the difference in area is negative.

(b) $$f : \left\{\begin{array}{l}\text{integers}\\ \text{greater than } 5\end{array}\right\} \longrightarrow \left\{\text{negative integers}\right\}$$

$$f(x) = -10x + 25$$



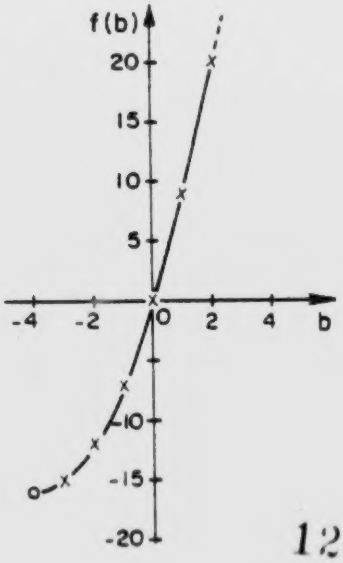
x	5	6	7	...
f(x)	-25	-35	-45	

III. (a) $$(4 + b)^2 - 16 = 16 + 8b + b^2 - 16 = 8b + b^2$$

$$f(b) = 8b + b^2 \ .$$

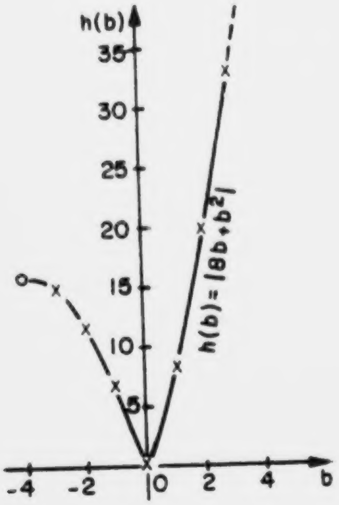
Here again we need to take into consideration the geometric limitations of the situation

i.e. $$f : \left\{\begin{array}{l}\text{real numbers}\\ \text{greater than } -4\end{array}\right\} \longrightarrow \left\{\begin{array}{l}\text{real numbers}\\ \text{greater than } -16\end{array}\right\}$$



(b) $$h : \left\{\begin{matrix}\text{real numbers}\\ \text{greater than } -4\end{matrix}\right\} \longrightarrow \left\{\begin{matrix}\text{positive}\\ \text{real number}\end{matrix}\right\}$$

$$h(b) = |8b + b^2| \ .$$



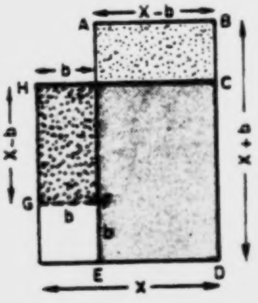
V. (a) The area of the rectangle is

$$(x + b)(x - b) = x^2 - b^2$$

The change in area is

$$x^2 - b^2 - x^2 = -b^2$$

This means that the change in area is independent of the side length and is a constant decrease of b^2 cm^2. This result seems somewhat strange, and it is a good idea to illustrate it graphically.



(b) This time the area of the rectangle is

$$(x + kx)(x - kx) = x^2 - k^2x^2 .$$

The change in area is $-k^2x^2$, which means a decrease proportional to the original area.

The geometric limitations of the problem imply that

$$x + kx > 0 \quad \text{and} \quad x - kx > 0$$

i.e. $x(1 + k) > 0$ and $x(1-k) > 0$.

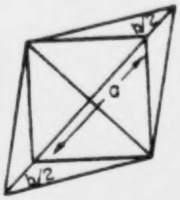
Since $x > 0$,

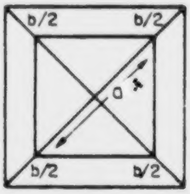
$$1 + k > 0 \quad \text{and} \quad 1 - k > 0$$

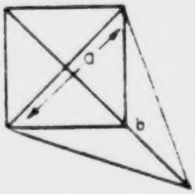
hence $|k| < 1$.

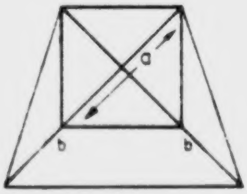
VI. There are many possible ways of extending the project. We mention two:

1. Consider a cube instead of a square and explore how changing the edges affects the volume.
2. Explore the effect of changing the diagonals of a square in various ways as illustrated (for example)





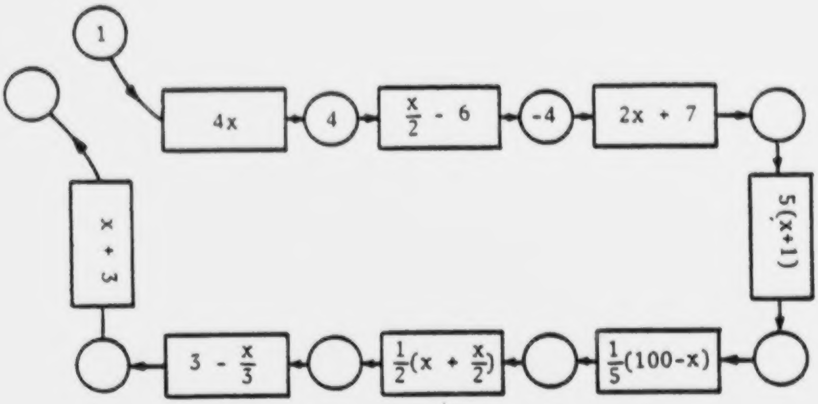




Assignment Project : Open phrases and magic circles

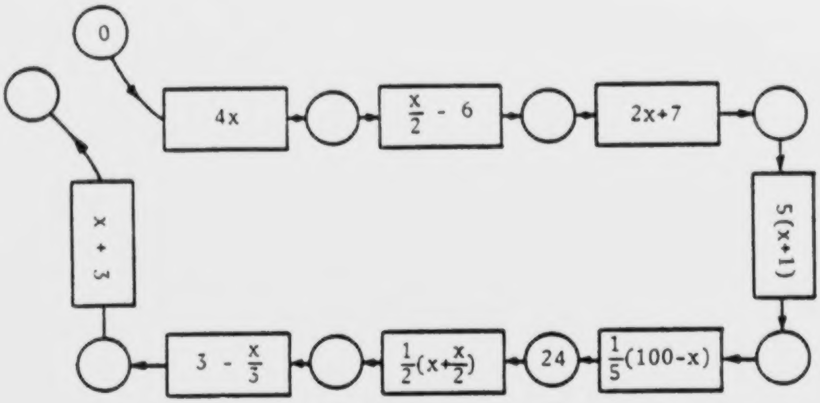
I. A circle of open phrases

Consider the following circle of open phrases:

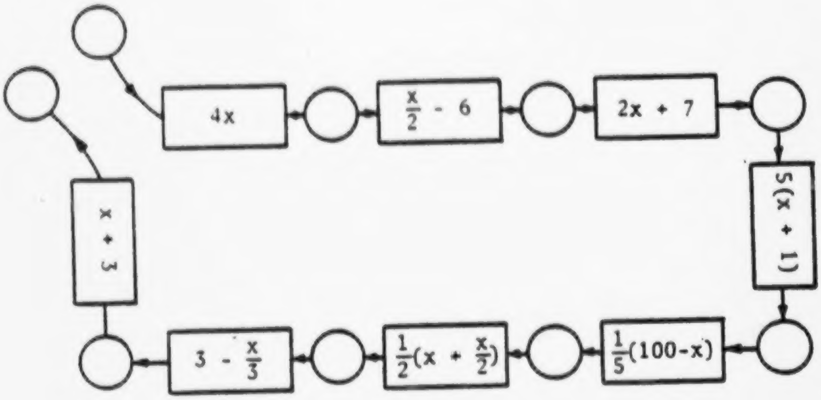

1
4x
4
$\frac{x}{2} - 6$
-4
2x + 7
5(x+1)
$\frac{1}{5}(100-x)$
$\frac{1}{2}(x + \frac{x}{2})$
$3 - \frac{x}{3}$
x + 3

(a) We begin by substituting 1 in 4x, the outcome 4 is recorded in the small circle to the right of the 4x box. We proceed clockwise and substitute the outcome in the next open phrase $\frac{x}{2} - 6$, and so on, until we come to the last one. What result is obtained from x + 3 ?

(b) Begin by substituting 0 in 4x and repeat the procedure.

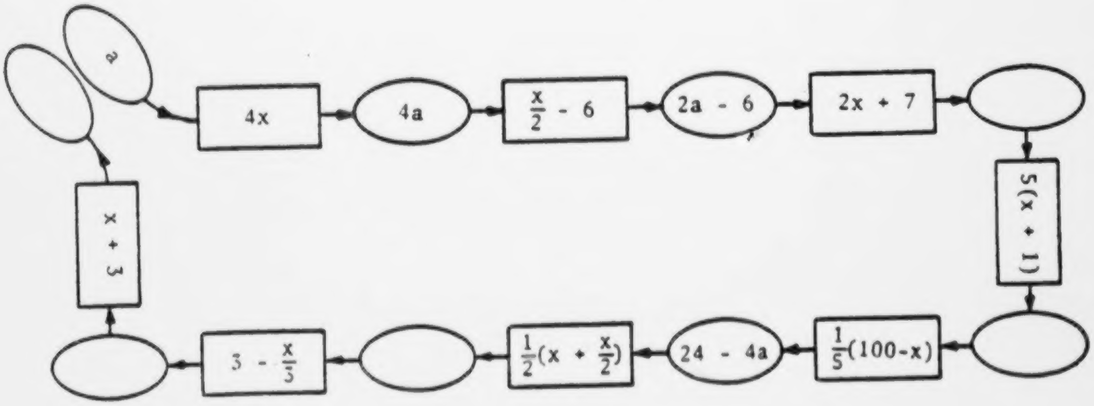

0
4x
$\frac{x}{2} - 6$
2x+7
5(x+1)
$\frac{1}{5}(100-x)$
24
$\frac{1}{2}(x+\frac{x}{2})$
$3 - \frac{x}{3}$
x + 3

(c) Choose any number to substitute in 4x and repeat the same procedure.


4x
$\frac{x}{2} - 6$
2x + 7
5(x + 1)
$\frac{1}{5}(100-x)$
$\frac{1}{2}(x + \frac{x}{2})$
$3 - \frac{x}{3}$
x + 3

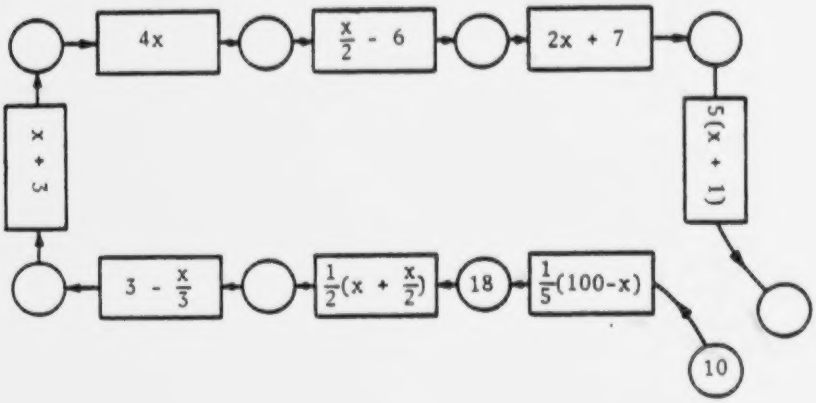
(d) Can you come to any conclusion?

(e) Substitute a variable (say, a) and follow it round the circle. What is the final outcome?

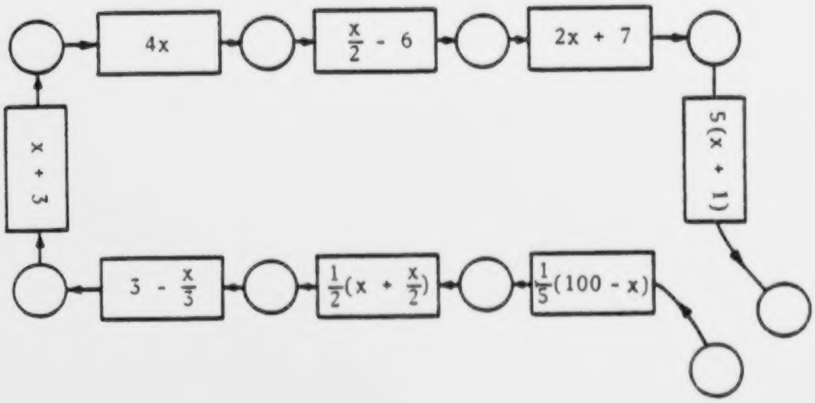

a
4x
4a
$\frac{x}{2} - 6$
2a - 6
2x + 7
5(x + 1)
$\frac{1}{5}(100-x)$
24 - 4a
$\frac{1}{2}(x + \frac{x}{2})$
$3 - \frac{x}{3}$
x + 3

II. Breaking the circle

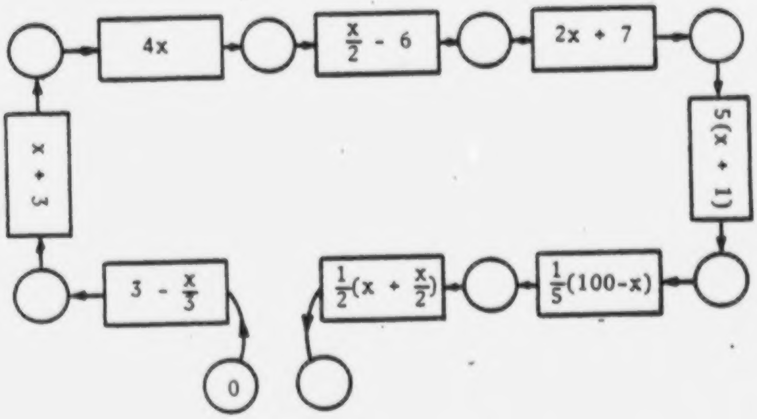
(a) Substitute 10 where we have broken the circle below, proceed clockwise and complete the circuit. What happens?


4x
$\frac{x}{2} - 6$
2x + 7
5(x + 1)
$\frac{1}{5}(100-x)$
18
$\frac{1}{2}(x + \frac{x}{2})$
$3 - \frac{x}{3}$
x + 3
10

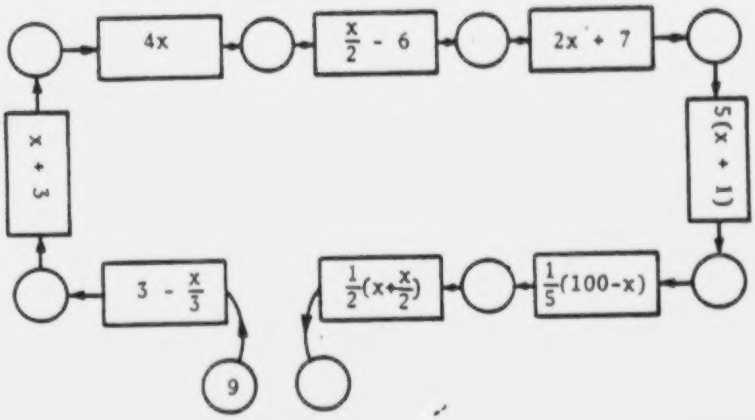
(b) Choose any number and repeat the procedure.


4x
$\frac{x}{2} - 6$
2x + 7
5(x + 1)
$\frac{1}{5}(100 - x)$
$\frac{1}{2}(x + \frac{x}{2})$
$3 - \frac{x}{3}$
x + 3

(c) What happens now?


4x
x/2 - 6
2x + 7
5(x + 1)
1/5(100-x)
1/2(x + x/2)
3 - x/3
x + 3
0

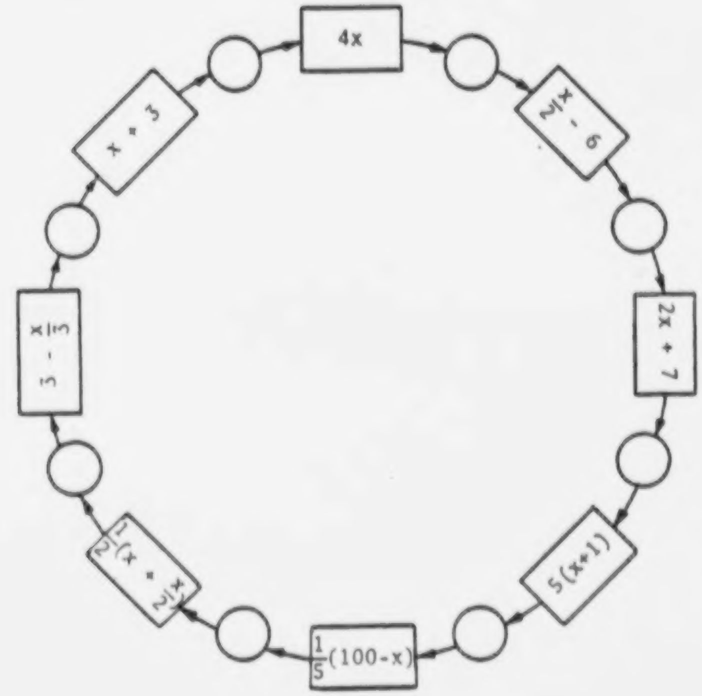
(d) Complete:


4x
x/2 - 6
2x + 7
5(x + 1)
1/5(100-x)
1/2(x+x/2)
3 - x/3
x + 3
9

(e) Break the circle at any point, substitute a number, proceed clockwise and complete the circuit.

(f) What can you say about the given circle?

III. A closed circle



(a) Which number was substituted in 4x if the outcome of $\frac{x}{2} - 6$ was 4?

(b) Which number was substituted in 4x if the outcome of 5(x + 1) was -60?

(c) Which number was substituted in 4x if the outcome of $3 - \frac{x}{3}$ was 9?

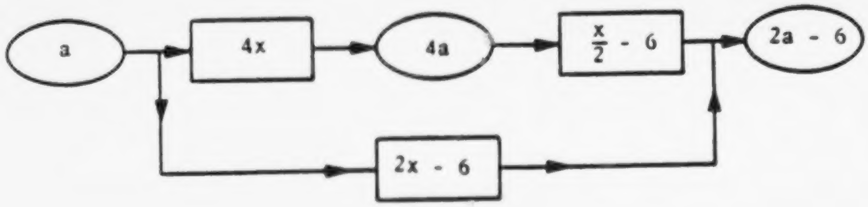
(d) The number 7 was substituted in 5(x + 1); what was the outcome of $3 - \frac{x}{3}$?

If we want the outcome of 2x + 7 to be 7, which number do we need to substitute in x + 3?

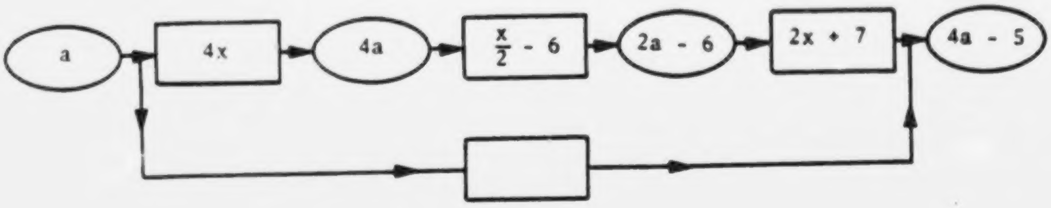
(e) The number 6 was substituted in $3 - \frac{x}{3}$; which was the outcome of $\frac{1}{5}(100 - x)$?

IV. Linear functions

The composition of the two linear functions $f(x) = 4x$ and $f(x) = \frac{x}{2} - 6$ is obtained as follows:



(a) What is the composition of the three following functions?



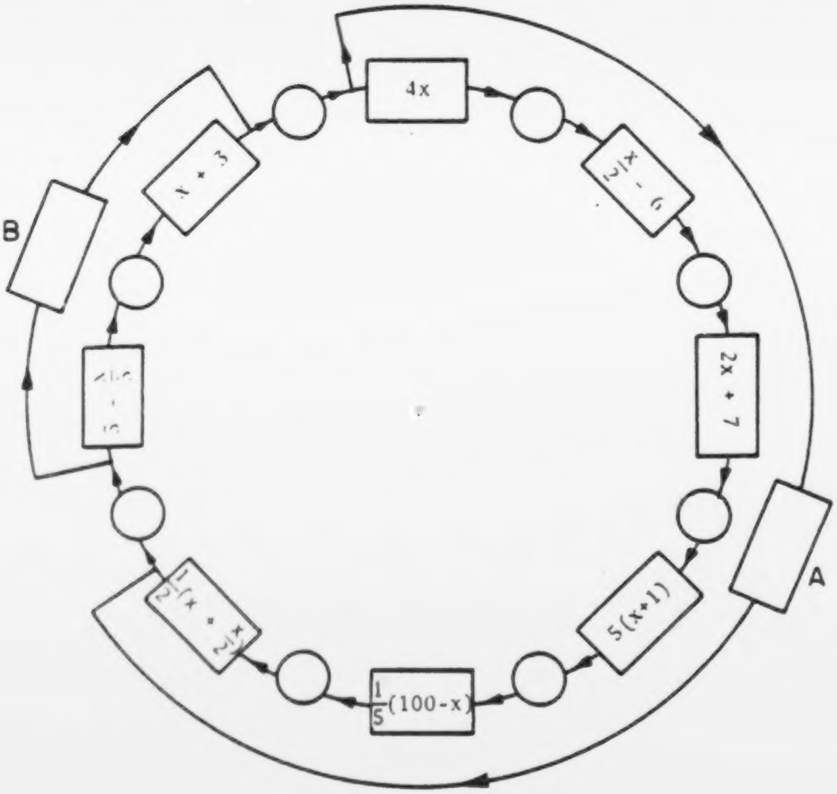
(b) Use your response to I-(e) to find the composition of the four linear functions starting with $f(x) = 4x$.

The same for the five, six, seven, eight functions?

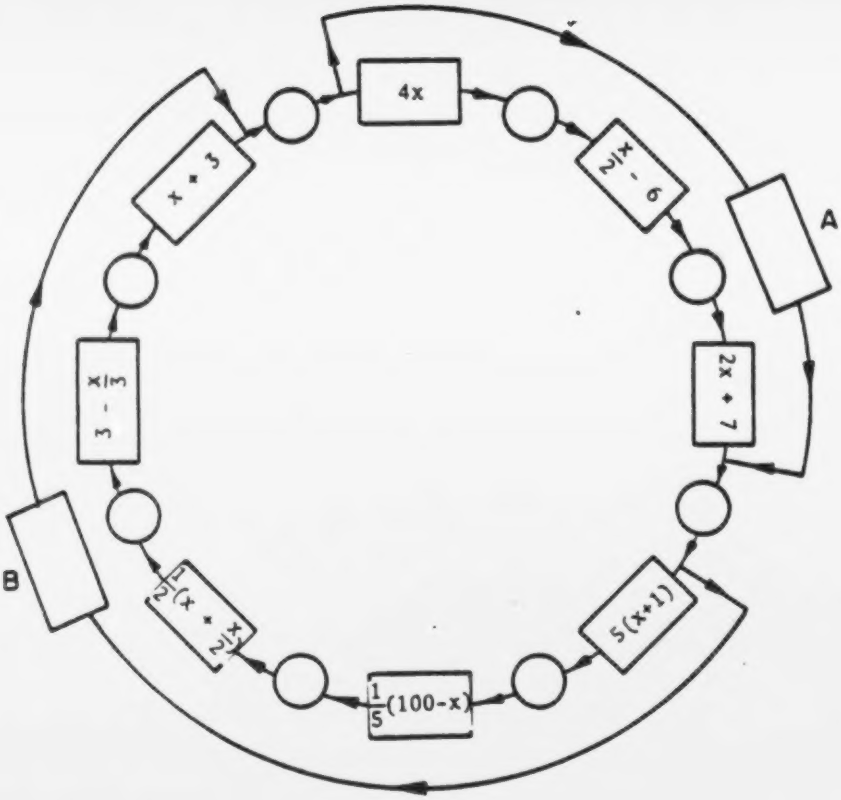
(c) Can you see a relation between the eighth function $f(x) = x + 3$, and the composition of the other seven functions?

(d) Find the two composite functions in each of the following figures and discuss their relation to each other in each case.

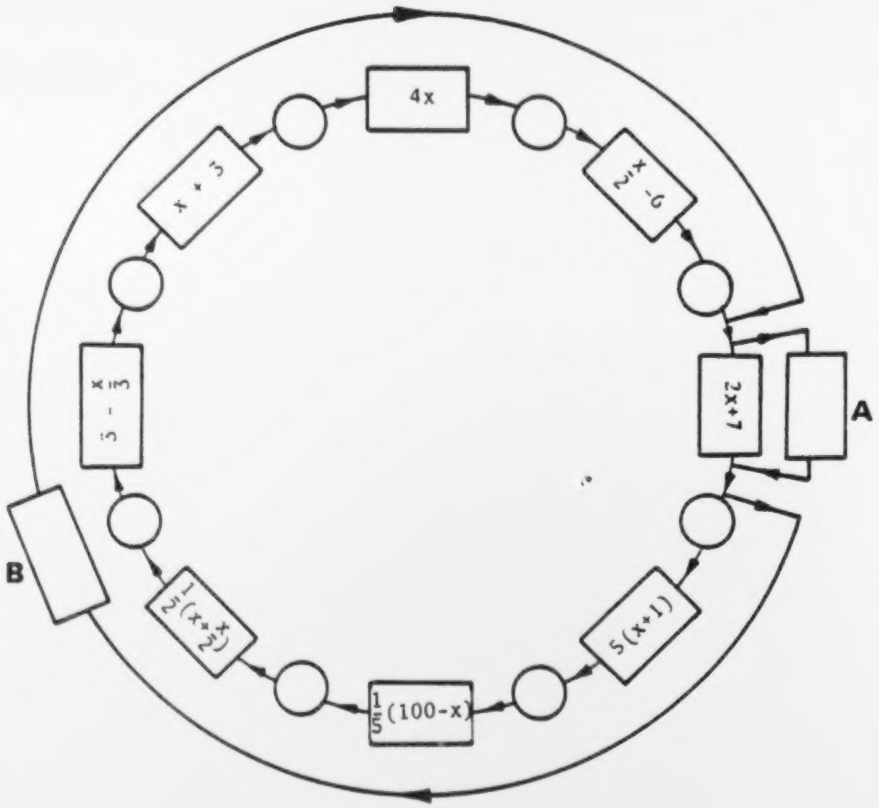
i)


4x
$\frac{x}{2} - 6$
2x + 7
A
5(x+1)
$\frac{1}{5}(100-x)$
$\frac{1}{2}(x + \frac{x}{2})$
B
x + 3

ii)

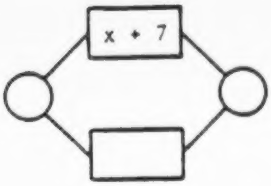

4x
$\frac{x}{2} - 6$
A
2x + 7
5(x+1)
$\frac{1}{5}(100-x)$
$\frac{1}{2}(x + \frac{x}{2})$
B
$3 - \frac{x}{3}$
x + 3

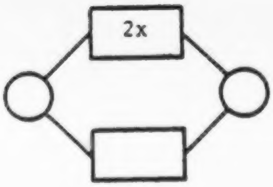
iii)

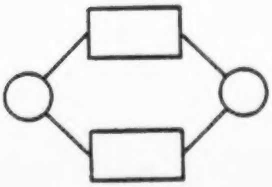

4x
$\frac{x}{2} - 6$
$2x+7$
A
$5(x+1)$
$\frac{1}{5}(100-x)$
$\frac{1}{2}(x+\frac{x}{2})$
B
$5 - \frac{x}{3}$
$x + 5$

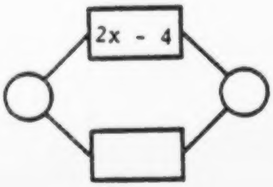
V. Constructing circles

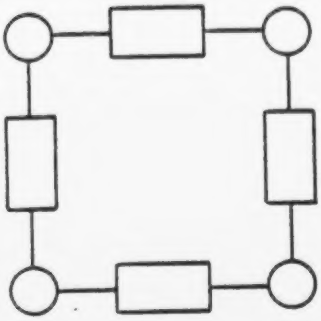
Construct closed circles which have the same properties as the circle we have considered so far.

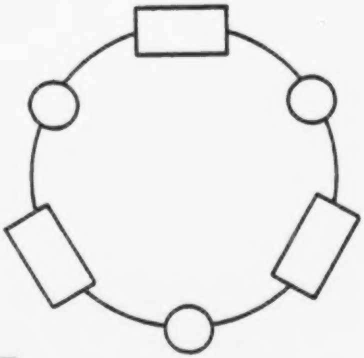


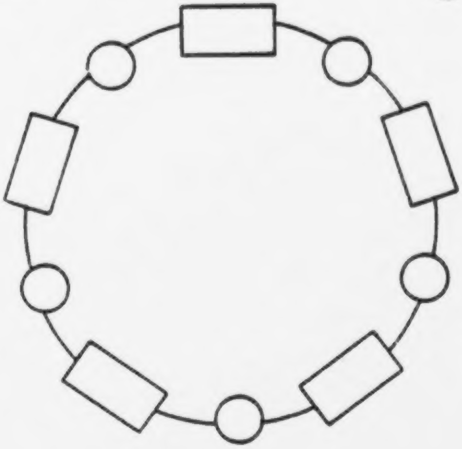










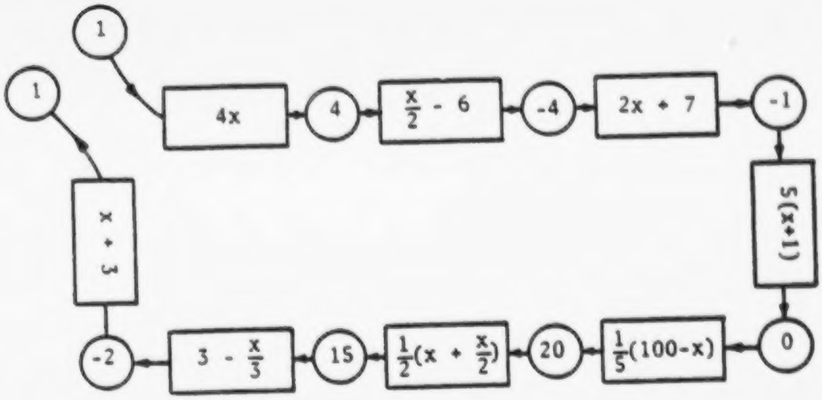


VI. Magic circles

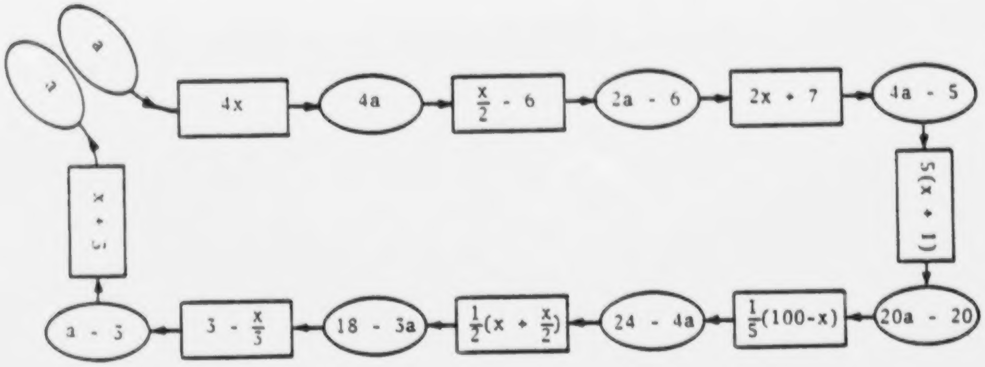
What happens if we go anticlockwise round the original circle? Is it possible to construct 'magic circles' such that, if we traverse them in either direction the output is the same as the input?

Solutions to selected sections

I (a) When we complete the circle the result obtained from x + 3 turns out to be 1, the number with which we started.


1
1
4x
4
x/2 - 6
-4
2x + 7
-1
5(x+1)
0
1/5(100-x)
20
1/2(x + x/2)
15
3 - x/3
-2
x + 3

(d) There are at least two ways of proving that the initial and final number are always the same. One way is to substitute a variable and follow it round the circle as in section (e).


a
a
4x
4a
x/2 - 6
2a - 6
2x + 7
4a - 5
5(x + 1)
20a - 20
1/5(100-x)
24 - 4a
1/2(x + x/2)
18 - 3a
3 - x/3
a - 3
x + 3

The second way of proving this result is more elegant but requires more mathematics. The whole circle reduces to a single linear function. Now a linear function is determined uniquely by two distinct pairs of numbers. Hence, if the substitution of two different numbers produces - at the end of the complete circle, the same two numbers respectively with which we started, we have proved that the circle reduces to the identity function.

II. (f) We can break it at any point, substitute a number, proceed clockwise, and complete the circuit at our point of entry with the number with which we started.

The mathematical justification:

We denote the functions in the circle as follows:

$$f_1(x) = 4x, \quad f_2(x) = \frac{x}{2} - 6, \ \ldots$$

As shown above, the composition of these functions in the correct order gives the identity function: that is

$$f_8 \circ f_7 \circ f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1 = I$$

(Note: function composition is associative, hence there is no need for brackets.) We wish now to prove that if we break the circle at any point, say between f_6 and f_7, and begin with f_7, we still get the identity function. That is, we wish to prove that

$$f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1 \circ f_8 \circ f_7 = I$$

Because function composition is associative, we may write

$$f_8 \circ f_7 \circ f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1 = (f_8 \circ f_7) \circ (f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1)$$

We denote the functions in the two brackets by F and G, respectively, and hence

$$F \circ G = I$$

A linear function is one-one and onto, and hence has an inverse function (which is also linear). Thus F and G are inverses, and we may also write

$$G \circ F = I$$

i.e.

$$(f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1) \circ (f_8 \circ f_7) = I$$

If we remove the brackets, we have what we wished to prove.

We can put this result in another (mathematical) way as follows:

If the composition of a sequence of linear functions is the identity function, then the composition of any cyclic permutation of the sequence is also the identity function.

III. (a) We "go back" two steps by solving equations.

Step 1: $\frac{x}{2} - 6 = 4$

$x = 20$

Step 2 : $4x = 20$

$x = 5$

(b) We are in the middle of the circuit and thus can obtain the answer in two ways.

(i) backwards

Step 1 : $5(x + 1) = -60$

$x = -13$

Step 2 : $2x + 7 = -13$

$x = -10$

Step 3 : $\frac{x}{2} - 6 = -10$

$x = -8$

Step 4 : $4x = -8$

$x = -2$

(ii) forwards

Step 1 : substitute - 60 in $\frac{1}{5}(100 - x)$ to get 32.

Step 2 : substitute 32 in $\frac{1}{2}[x + \frac{x}{2}]$ to get 24.

Step 3 : substitute 24 in $3 - \frac{x}{3}$ to get - 5.

Step 4 : substitute -5 in $x + 3$ to get -2.

(c) We can go forward one step, substitute 9 in $x + 3$ and get 12.

(d) The answer to both questions is 0. This is a hint to the idea of inverse functions, which is needed in the following part.

(e) $\frac{1}{2}(x + \frac{x}{2}) = 6$

$x = 8$

IV.

(c) The composition of the seven functions is $f(x) = x - 3$, which is the inverse of $f(x) = x + 3$.

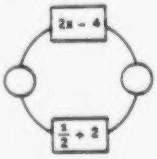
(d) We get pairs of inverse functions:

(i) $f(x) = 18 - 3x$ $f(x) = -\frac{x}{3} + 6$,

(ii) $f(x) = 4x - 5$ $f(x) = \frac{x}{4} + \frac{5}{4}$,

(iii) $f(x) = 2x + 7$ $f(x) = \frac{x}{2} - \frac{7}{2}$.

V. Two inverse functions give a two-function circle:



If we want a circle of more than two functions we can use pairs of inverse functions in sequence, but this is very dull.

More general and efficient methods for building circles can be based on the proofs we gave above related to part II.

(i) If we want an n-function circle, we can choose any n - 1 linear functions, find their composition, and then the nth function is the inverse of this composition. The only drawback to this method is that composition of functions (even if they are linear) is somewhat time-consuming if n is not small.

(ii) A somewhat less laborious method is to choose any n - 1 linear functions and then 'pass through' them any two numbers, for example 0 and 1. Assume that the 'output' is p and q, respectively. Now all we have to do is to find the linear function f_n for which

$$f_n(p) = 0, \quad f_n(q) = 1 .$$

Example: To build a 3-function circle, whose first two functions are

$$f_1(x) = 2x + 3, \quad f_2(x) = 7 - x$$

We find the images of 0 and 1:

$$0 \to \boxed{2x+3} \to 3 \to \boxed{7-x} \to 4$$

$$1 \to \boxed{2x+3} \to 5 \to \boxed{7-x} \to 2 .$$

Let $f_3(x) = ax + b$ then we have to find a and b such that

$$4a + b = 0$$

$$2a + b = 1 ,$$

i.e. $f_3(x) = -1/2x + 2$.

VI. It is possible to construct "magic circles" such that if we traverse them in either direction the output is the same as the input.

We answer the question at three (ascending) levels of sophistication.

(a) Trivially, we can form a 2n-function magic circle by taking linear functions and their inverse: thus

$$f^{-1} \circ f \circ g^{-1} \circ g\, h^{-1} \circ h = h \circ h^{-1} \circ g \circ g^{-1} \circ f \circ f^{-1} = I$$

(b) If we can find a set of n mutually commutative linear functions, whose composition is the identity, then we have an n-function magic circle. We, therefore, have first to find the condition that two linear functions

$$f(x) = ax + b \qquad a \neq 0$$

$$g(x) = cx + d \qquad c \neq 0$$

commute.

$$(f \circ g)(x) = acx + ad + b$$

$$(g \circ f)(x) = acx + cb + d$$

Hence fog = gof if and only if

$$ad + b = cb + d$$

or $$d(1 - a) = b(1 - c)$$

If a, c ≠ 1, then we can rewrite this condition such that the parameters of f and g are separated:

$$\frac{b}{1 - a} = \frac{d}{1 - c} .$$

What is the meaning of $b/(1 - a)$? It is the value of the fixed point of f; i.e. the intersection of the graph of f with the graph of the identity function $I(x) = x$. For, the fixed point is the solution of the equation

$$x = ax + b$$

which, when $a \neq 1$, gives $x = b/(1 - a)$.

Similarly $d/(1 - c)$ gives the fixed point of g.

Thus, if a, c ≠ 1, we may conclude that fog = gof, if and only if f and g have the same fixed point (or, their graph intersect on $y = x$). There are some special cases when a and/or c is equal to 1. Thus

$$a = c = 1 \Rightarrow d(1 - a) = b(1 - c) .$$

Geometrically, f and g are then represented by lines parallel to $y = x$. If only one of a or c is equal to 1, then the corresponding function is the identity and hence the commutative condition is trivially satisfied. We can now summarize our findings. We can form a magic circle from a set of functions whose composition is I and either

(i) the slope of each of the functions is 1,

or

(ii) all the functions have the same fixed point.

(c) The methods described in (a) and (b) were based on sufficient conditions for a circle to be magic, but they are clearly not necessary. In this section we look for a necessary and sufficient condition, i.e. the general magic circle of linear functions.

We require an 'if and only if' condition equality

$$f_n \circ f_{n-1} \circ \ldots \circ f_1 = f_1 \circ f_2 \circ \ldots \circ f_n = I .$$

We write

$$f_1 \circ f_2 \circ \ldots \circ f_{n-1} = G$$

$$f_{n-1} \circ \ldots \circ f_2 \circ f_1 = H$$

and our equality becomes

$$f_n \circ H = G \circ f_n = I .$$

This equality holds if and only if f_n is the inverse of both G and H. But the inverse is unique; therefore, we must have

$$H = G$$

To satisfy this latter condition, we write

$$f_1 \circ f_2 \circ \ldots \circ f_{n-2} = G'$$

$$f_{n-2} \circ \ldots \circ f_2 \circ f_1 = H'$$

and the condition becomes

$$f_{n-1} \circ H' = G' \circ f_{n-1}$$

Now H' and G' are the composition of the same linear functions, hence they have the same slope (the product of the slopes of $f_1, \ldots, f_{n-2}$). Hence, we may write

$$G'(x) = Ax + B_1$$

$$H'(x) = Ax + B_2$$

$$f_{n-1}(x) = cx + d ,$$

whence $f_{n-1} \circ H' = G' \circ f_{n-1}$ if and only if

$$Ad + B_1 = cB_2 + d ,$$

or

$$d(1 - A) = B_1 - cB_2 .$$

We can always find c, d to satisfy this equation, unless $A = 1$, $B_1 \neq 0$, $B_2 = 0$.

We can now describe how to construct a general n-function magic circle:

(i) Choose any n-2 linear functions and find the corresponding A, B_1, and B_2. (An efficient way of doing this is to find A as the product of the slopes and B_1 and B_2 by passing 0 through the n-2 functions in the appropriate order. In the unfortunate event that $A = 1$, $B_1 \neq 0$ and $B_2 = 0$, change one of the functions.)

(ii) Find $f_{n-1}(x) = cx + d$ by choosing any c and d which satisfy the equation

$$d(1 - A) = B_1 - cB_2$$

(Note that if $A = 1$, d is arbitrary and c is determined uniquely. Otherwise, in general, one of c or d may be chosen arbitrarily and the other is then determined.)

(iii) Find f_n, the inverse of the composition of $f_1, \ldots, f_{n-1}$.

Example: A 4-function magic circle.

We choose $f_1(x) = 2x + 3$, $f_2(x) = 8 - x$.

Hence $A = -2$, $B_1 = 19$, $B_2 = 5$.

Any c and d which satisfy

$$3d = 19 - 5c$$

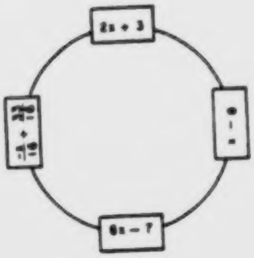
will define an appropriate f_3. We chose $f_3(x) = 8x - 7$.

The image of 0 through f_1, f_2 and f_3 is 33.

The image of 1 through f_1, f_2 and f_3 is 17.

Hence $f_4(x) = -x/16 + 33/16$.

and we have the magic circle:



Assignment Project: Elections

I. A ninth grade class (32 students) held elections to choose their representative on the school student council. The candidates were Dan and Eyal, who were not allowed to vote.

(a) Denote by x the votes for Dan and by y the votes for Eyal. The result of the election will be represented by the order pair (x, y).

Who was elected and how many abstained if the result was

(i) (21,7), (ii) (12,18), (iii) (13,13), (iv) (13,23).

Write down and discuss some other pairs which can represent the result of the election and give also examples of pairs which cannot.

(b) Illustrate graphically the fact that the number of voters does not exceed 30.

On your drawing mark a point

(i) A-which corresponds to Dan being elected.

(ii) B-which corresponds to Eyal being elected and no abstentions.

(iii) C-which corresponds to a tie.

(iv) D-which does not correspond to any possible result of the election.

(c) Dan has 4 close friends in the class, who certainly vote for him. Eyal has 9 sure votes. Illustrate these facts in your drawing and use it to:

(i) find three different possible results in which Dan is elected, and three in which Eyal is elected.

(ii) mark all the points which correspond to a tie. What do you notice?

(iii) decide who has more chance of winning.

(iv) find how many possible results of the election there are. In how many of them does Dan win and in how many Eyal?

(v) find what are the percentage chances of winning for each of the two candidates.

(d) Suppose that we know that Eyal was elected and that he polled more than 3 times the number of votes for Dan.

Write open-sentences to describe all the information you have.

Illustrate it graphically in another drawing and use it to:

(i) mark three possible results of the election.

(ii) find the total number of possible results.

(iii) find the result of the election in which there were the minimum number of voters.

II. Another ninth grade class (36 students) also held elections. The candidates where Nir and Mike and they were allowed to vote.

(a) Denote by x the votes for Nir and by y the votes for Mike. For each of the ten following pairs, find the percentage of voters who did not abstain.

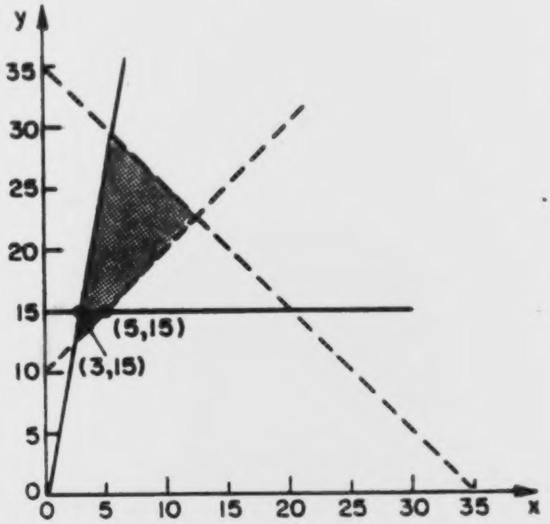
(14,13) (11,7) (5, 4) (20, 7) (12,28)

(9, 9) (25,7) (20,19) (23,10) (25,11)

(b) If we know that the percentage voting was at least 50% and that Nir got more than 10 votes but less than 18 more than Mike, write open-sentences which describe all the information you have and illustrate graphically.

(c) How many different possible results of the election are there in (b)?

III. The possible outcomes of a third ninth grade election are illustrated in the following drawing:



What can you say about the election?

IV. Describe verbally, algebraically and graphically possible elections in a fourth class.

Solutions to selected sections

I. (b) The shaded area illustrates the truth set of $x + y \leqslant 30$ and $x \geqslant 0$ and $y \geqslant 0$

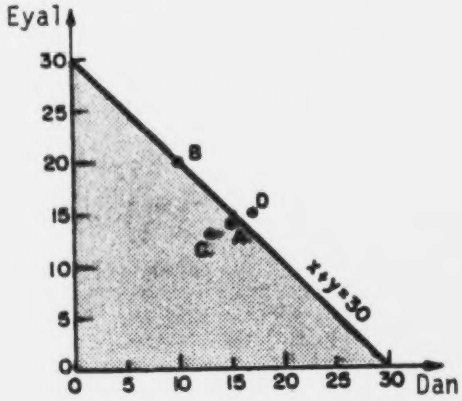
For example

(i) A(15,14)

(ii) B(10,20)

(iii) C(13,13)

(iv) D(17,15).



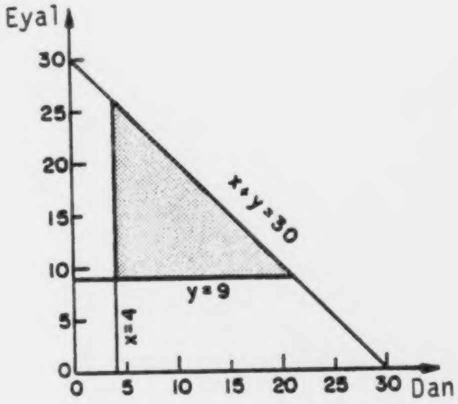
(c) The shaded area illustrates the truth set of $x + y \leqslant 30$ and $x \geqslant 4$ and $y \geqslant 9$

(i) Examples in which Dan is elected are

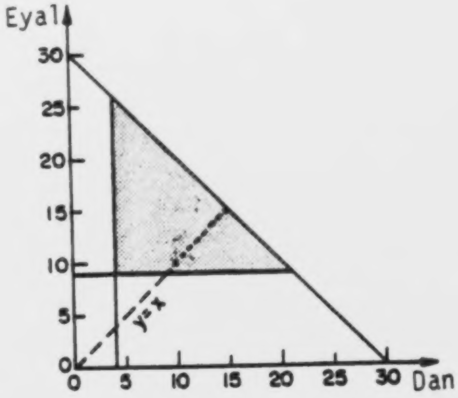
(14,13) (15,11) (19,11).

Examples in which Eyal is elected are

(12,13) (6,22) (8,22)



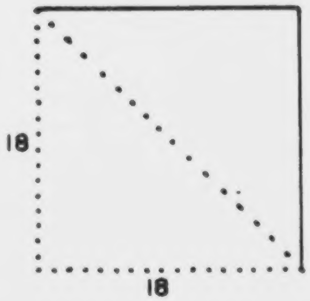
(ii) The points which correspond to a tie are those points which have integer coordinates and lie on that part of the line $y = x$ within the shaded area.



(iii) One can see from the drawing that the area for which $y > x$ is greater than the area for which $y < x$. This means that Eyal has a better chance of winning. This is clearly due to the fact that Eyal has more sure votes.

(iv) There are 171 possible results.

If the students use graph paper for the drawing they can obtain the answer by counting points. The answer can also be obtained by more sophisticated methods, for example:



(1) Complete the shaded triangle to a square.

There are 18 lines, each with 18 points.

18 points lie on the diagonal and $\frac{17 \times 18}{2}$ below the diagonal.

Hence the number of possible results is

$\frac{17 \times 18}{2} + 18 = 171$.

(2) We need to find sum of all integer numbers from 1 to 18. The numbers can be added in pairs:

1+18, 2+17, etc. There are 9 pairs, and hence $19 \cdot 9 = 171$ possible outcomes.

(v) Eyal can win in 122 cases and Dan in 42 cases. In 7 cases a tie will occur.

In percentages:

$\frac{122 \cdot 100}{171} \cong 71.3\%$ $\frac{42 \cdot 100}{171} \cong 24.5\%$

(d) The shaded area illustrates the truth set of

$x + y \leqslant 30$ and $x \geqslant 4$

and $y \geqslant 9$ and $y > 3x$

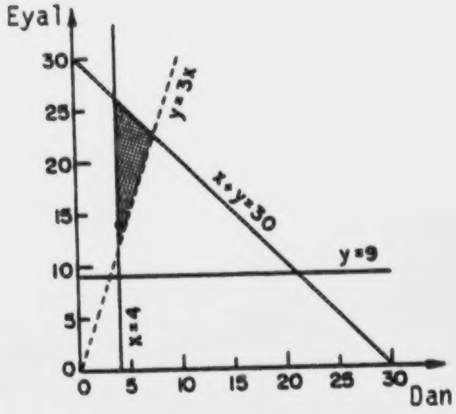
(i) for example:

(6,22) (6,24) (5,20).

(ii) There are 32 possibilities.

(iii) The result with a minimum number of voters is (4,13).

The point is marked in the drawing.

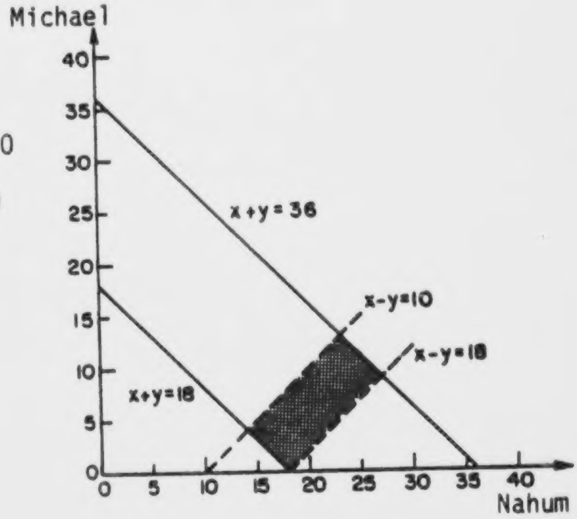


II. (b) The shaded area illustrates the truth set of

$x + y \leqslant 36$ and $x \geqslant 0$ and $y \geqslant 0$

and $x + y \geqslant 18$ and $x - y > 10$

and $x - y < 18$.



III. The related open sentences are:

$x + y < 35$ and $y > 15$ and $y > x + 10$ and $y < 5x$.

Some possible conclusions are:

(i) Not more than 35 children were eligible to vote.

(ii) The percentage voting is greater than 50%.

(iii) The candidate whose votes are denoted by y was elected and got more than 15 votes.

(iv) The winner got at least 11 votes, but not more than 22, more than the loser.

APPENDIX B

EVALUATION INSTRUMENTS

Mathematics Achievement Test (ACH)

Functions

Note: Unless specified, the domain and range are the real numbers.

1. Given an onto function $f : x \longmapsto 2x^2 - 1$ with domain

   $\{-3, -2, -1, 0, 1, 2, 3\}$.

   The range is

   (a) $\{\text{integers}\}$ (b) $\{-1, 1, 7, 17\}$

   (c) $\{0, 1, 7, 17\}$ (d) $\{1, 7, 17\}$.

2. Given the function $f: x \longmapsto 2x + 5$,

   The pre-image of -1 is

   (a) -2 (b) 2

   (c) -3 (d) 3

3. Given the function $g(x) = 2x - |x|$, the value of $g(-2)$ is

   (a) -6 (b) 2

   (c) -2 (d) 6

4. The image-set of the function $y = 3x + 12$ is $\{y/y > 0\}$.

   Its domain is

   (a) $\{x|\ x > 0\}$ (b) $\{x|\ x > 4\}$

   (c) $\{x|\ x > -4\}$ (d) $\{x|\ x < -4\}$

5. The domain of the function $y = -2x + 6$ is

   $x|x > 0$. Its image-set is

   (a) $\{y|y > o\}$ (b) $\{y|y < 3\}$

   (c) $\{y|y > 3\}$ (d) $\{y|y < 6\}$

6. Given the two functions $f(x) = x^2$, $g(x) = 5x + 3$

The value of $g(f(3))$ is

(a) 324 (b) 48

(c) 162 (d) 228

7. Given the two functions: $f(x) = x^2$, $g(x) = 5x + 3$

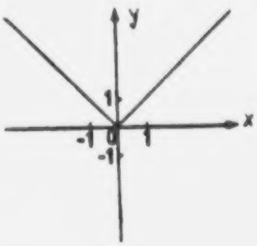
The function $g(f(x))$ is

(a) $(5x + 3)^2$ (b) $5x^2 + 3$

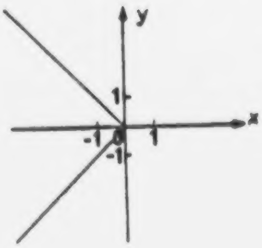
(c) $(5x)^2 + 3$ (d) $(5x + 3)\cdot x^2$

8. The graph of $y = -|x|$ is

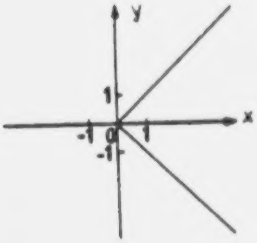
(a)



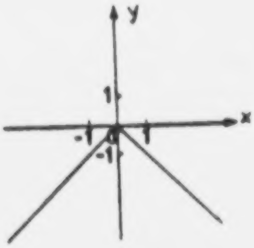
(b)



(c)



(d)



9. Given the function $f(x) = |x + 1|$, the pre-images of 2 are

(a) 2 and -2 (b) 1 and -2

(c) 1 and -1 (d) 1 and -3

10. The function $f(x) = x + 1$ decreases in the domain:

(a) $\{x|x < 1\}$ (b) $\{x|x > 1\}$

(c) $\{x|x < -1\}$ (d) $\{x|x < -1\}$

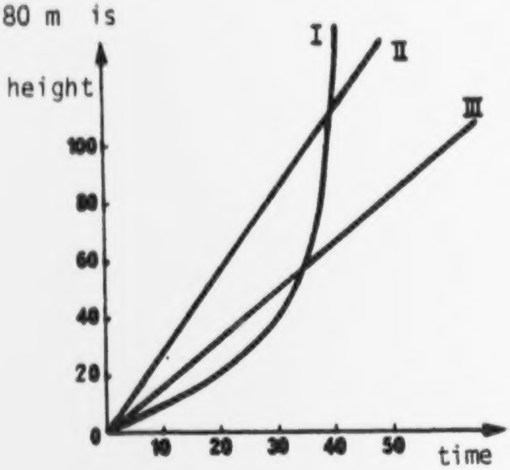
11. The graphs illustrate the movement of three objects. The order in which they reach a height of 80 m is

(a) first III, then II, then I

(b) first I, then II, then III

(c) first II, then III, then I

(d) first II, then I, then III



12. The points with coordinates (-3, 0), (2, 0) and (0, 0) are the intersections with x-axis of the function

(a) $y = x(x - 2)(x + 3)$ (b) $y = (x-2)(x + 3)$

(c) $y = x(x + 2)(x - 3)$ (d) $y = (x+2)(x - 3)$

13. Consider the family of functions $y = 2x + b$, where b is any real number. Which of the following sentences is true?

Among the graphs of the functions in the family there is

(a) a pair which meet on the x-axis.

(b) a pair which meet on the y-axis.

(c) a graph parallel to the x-axis.

(d) a graph which passes through the origin.

14. The price of a product is raised by 20% and then again by 20%. The final price is represented by

(a) $y = \frac{x \cdot 20 \cdot 20}{100}$ (b) $y = \frac{x \cdot 120 \cdot 120}{100 \cdot 100}$

(c) $y = \frac{x \cdot 120 \cdot 120}{100}$ (d) $y = \frac{x \cdot 20 \cdot 20}{100 \cdot 100}$

15. To convert temperature from Fahrenheit to Centigrade, one first subtracts 32° and then multiplies by $\frac{5}{9}$.

The fraction which relates Fahrenheit to Centigrade is

(a) $t \longmapsto \frac{5}{9}t - 32$ (b) $t \longmapsto t - 32 \cdot \frac{5}{9}$

(c) $t \longmapsto \frac{5}{9}(t+32)$ (d) $t \longmapsto \frac{5}{9}(t-32)$

16. The function $y = 5x - 4$ intersects the axes in the points

(a) (0, -4) , (0, 5) (b) (-4, 0) , (0, 5)

(c) (0.8, 0) , (0, -4) (d) (-0.8, 0) , (0, 4).

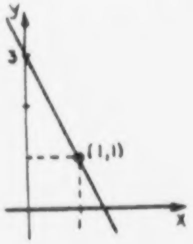
17. The function with graph as shown in the figure is

(a) $y = -\frac{1}{2}x + 3$

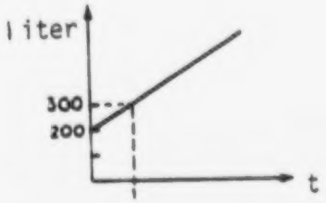
(b) $y = 2x + 3$

(c) $y = \frac{1}{2}x + 3$

(d) $y = -2x + 3$



18. The graph representing the function which relates time with the amount of water in a container is



(a) $y = 100x + 200$ (b) $y = 100x + 300$

(c) $y = 300x + 200$ (d) $y = 200x + 100$

19. Consider the family of functions $y = ax + 3$, where a is any real number.

Which of the following sentences is true?

Among the graphs of the function there is

(a) a pair of parallel graphs.

(b) a pair which meet on the x-axis.

(c) a graph which passes through the origin.

(d) a graph parallel to the x-axis.

20. The graph of a function passes through A(3, 5) and is parallel to the line $y = -2x + 4$.

The function is

(a) $y = -2x + 5$ (b) $y = -2x + 11$

(c) $y = \frac{7}{3}x - 2$ (d) $y = \frac{x}{3} + 4$.

Attitude questionnaire

This questionnaire is not a test. Its object is to learn about your attitude toward studying mathematics.

For each statement circle the number which closely expresses your attitude.

	Strongly agree	Agree	Tend to agree	Tend to disagree	Disagree	Strongly disagree
1. I find mathematics easy.	6	5	4	3	2	1
2. I dislike mathematical questions which require exploring.	6	5	4	3	2	1
3. Students should not be expected to answer mathematics questions on topics they have not learned.	6	5	4	3	2	1
4. I am not that successful in mathematics.	6	5	4	3	2	1
5. I solve mathematical problems only if they are given as homework.	6	5	4	3	2	1
6. I often have trouble mastering the material we study in mathematics.	6	5	4	3	2	1
7. In mathematics one does not have to pose questions, only to find answers.	6	5	4	3	2	1
8. I feel that I usually master the material we study in mathematics.	6	5	4	3	2	1
9. I find it interesting to prove things in mathematics.	6	5	4	3	2	1
10. I prefer easy questions.	6	5	4	3	2	1
11. I usually need help from the teacher to solve mathematical problems.	6	5	4	3	2	1
12. Students can pose interesting questions relevant to the material they study in mathematics.	6	5	4	3	2	1
13. I would like to read extra material in mathematics.	6	5	4	3	2	1
14. I remember most of the things I have learned in mathematics.	6	5	4	3	2	1
15. Students should justify their responses to mathematical problems.	6	5	4	3	2	1

	Strongly agree	Agree	Tend to agree	Tend to disagree	Disagree	Strongly disagree
16. I can solve mathematical problems easily.	6	5	4	3	2	1
17. I find mathematics interesting and am curious to learn more.	6	5	4	3	2	1
18. I like to work on new material.	6	5	4	3	2	1
19. Sometime one can answer mathematical questions,even when they are related to unfamiliar material.	6	5	4	3	2	1
20. When a problem looks difficult to me I tend to give up.	6	5	4	3	2	1
21. I try to find alternative solutions to mathematical problems.	6	5	4	3	2	1
22. I sometimes read chapters in the mathematics text before we study them.	6	5	4	3	2	1

Evaluation schemes for the assignments 1 and 3

Pretest

Assignment 1 : Which numbers when substituted in $3x + 1$, will give even numbers?

Answer scale

A_1 some odd numbers

11 some examples of odd numbers,

12 other (confusion of odd and even).

A_2 the pattern "all odd numbers"

21 generalization for odd numbers,

22 generalization for odd numbers and numerical examples such as $\frac{1}{3}$,

23 other.

A_3 all numbers of the form $\frac{2n - 1}{3}$ (n an integer)

31 all numbers of the form $\frac{2n - 1}{3}$

32 all odd numbers, all odd numbers $+ \frac{2}{3}$, and all even number $+ \frac{1}{3}$

33 other.

Procedure scale

P_1 guess and check procedure

11 no explanation,

12 substitution of odd numbers only,

13 substitution of odd and even numbers,

14 also substitution of fraction,

15 other.

P_2 proof of one condition (sufficient or necessary)

21 incomplete proof of sufficient condition; relate only to +1 or only to 3x,

22 proof of sufficient condition: if x is odd then $3x + 1$ is even,

23 incomplete proof of necessary condition:

$$3x = 2a - 1$$

$$\text{or} \quad 3x + 1 = 2x \qquad x = -1$$

24 proof of necessary condition: an even number minus 1 is an odd number; an odd number divided by 3 without a remainder must be odd,

25 other.

P_3 sufficient and necessary proof

31 proof that if x is odd then $3x + 1$ is even and if x is even $3x + 1$ is odd,

32

$$3x + 1 = 2n$$

$$x = \frac{2n - 1}{3},$$

33 other.

Post-test

Assignment 1 : Which numbers when substituted in $x^2 - 5x$, give negative numbers?

Answer scale

A_1 wrong answer

11 $x > 5$,

12 $x > 0$,

13 other.

A_2 partial answer

21 $x < 5$,

22 $x > 0$,

23 $-5 < x < 5$,

24 relating to integers only ,

25 other.

A_3 full answer

31 $0 < x < 5$

Procedure scale

P_1 checking numerical examples

11 no explanation,

12 check of numerical examples,

13 check of $x = 0$ and/or $x = 5$ only,

14 other.

P_2 incomplete justification

21 consideration of $x > 5$ and $x < 5$ only,

22 consideration of $x < 0$ and $x > 0$ only.

23 attempt at necessary condition from $x(x - 5)$,

24 other.

P_3 complete justification

31 consideration of both $x = 0$ and $x = 5$,

32 solution of inequality $x^2 - 5x < 0$

33 use of the graph of $y = x^2 - 5x$,

34 32 with a mistake $x < 5$ and adding a correction as 31.

pretest

Assignment 3 : Given the equation $b(2x + b) = 175$ where x and b represent positive integers, what can you say about x and b ?

Answer scale

A_1 (at most) one value for b and one value for x

11 no conclusion

12 b is an odd number,

13 b   13 ,

14 $b = 5 \quad x = 15$,

15 $b = 1 \quad x = 87$

16 $b = 7 \quad x = 9$,

17 other.

A_2 two answers

21 $b = 5 \quad x = 15; \quad b = 1 \quad x = 87$,

22 $b = 5 \quad x = 15; \quad b = 7 \quad x = 9$,

23 $b = 1 \quad x = 87; \quad b = 7 \quad x = 9$,

24 other (negative numbers also).

A_3 the full answer

31 $b = 1 \quad x = 87; \quad b = 5 \quad x = 15;$
$b = 7 \quad x = 9$,

32 other.

Procedure scale

P_1 numerical substitution and/or algebraic manipulation

11 no explanation or just repetition of the data,

12 substitution for one variable and solution for the other,

12 substitution for both variables in $b(2x + b)$,

14 an attempt to express one variable in terms of the other,

15 other.

P_2 some general conclusion

21 $x = \dfrac{175 - b^2}{2b}$,

22 b divides 175,

23 investigation of the sum $2bx + b^2 = 175$,

24 other.

P_3 as in P_2 together with the fact that b and x are positive integers

31 systematic substitution for one variable to exhaust all possibilities,

32 systematic check of all factors of 175,

33 other.

posttest

Assignment 3 : Given the system of open sentenses

$$50 + 2a + b = 7b$$

$$50 + 2a + b < 100$$

where a and b represent positive integers, what can you say about a and b?

Answer scale

A_1 (at most) one value for a and one value for b

11 no conclusion,

12 $7b < 100$,

13 $a = 5 \quad b = 10$,

14 another pair,

15 other.

A_2 several answers

21 some pairs without the end pairs,

22 only the end pairs,

23 $b < 14\frac{2}{7}$ and $a < 17\frac{6}{7}$,

24 a from 1 to 17 and
b from 1 to 14 ,

25 other.

A_3 the full (or almost full) answer

31

a	2	5	8	11	14	17
b	9	10	11	12	13	14

32 other.

Procedure scale

P_1 numerical and/or algebraic manipulation

11 relation between a and b but no connection between the open sentences,

12 connection between the open sentences but no relation between a and b,

13 other.

P_2 using part of the data

21 connection between the open sentences to find pairs (a, b),

22 connection between the open sentences and use of the fact that a, b are integers,

23 connection between the open sentences and use of the fact that a, b are positive,

P_3 Use of all the data

31 connection between the two open sentences and use of the fact that a, b are positive integers,

32 other.

APPENDIX C

COMPARATIVE STUDY :

GROUP COMPARISON OF AVERAGE SCORES AND DISTRIBUTION TABLES

	Group / Test	Experimental n = 191		Control n = 194		t	p
		$\overline{x}$	σ	$\overline{x}$	σ		
	ACH	59.2	(19.1)	57.4	(20.3)	0.89	n.s.
	ATSI	68.2	(13.8)	67.0	(13.4)	0.85	n.s.
	ATMA	67.8	(12.6)	67.9	(11.4)	-0.12	n.s.
	Q	8.78	(2.15)	8.67	(2.00)	0.52	n.s.
assignment 1-3 maximum score: 6	Q_1	3.55	(0.77)	3.58	(0.77)	-0.41	n.s.
	Q_2	2.96	(1.18)	3.07	(1.21)	-0.94	n.s.
	Q_3	2.27	(1.19)	2.01	(1.08)	2.21	0.03

Table C_1 : Group comparison of average scores - pretest

Group / Test	Experimental n = 191		Control n = 194		t	p
	$\bar{x}$	σ	$\bar{x}$	σ		
ACH	66.6	(20.7)	58.1	(19.9)	4.09	0.0001
ATSI	67.2	(14.2)	64.6	(13.6)	1.86	0.06
ATMA	68.4	(12.9)	65.5	(12.0)	2.26	0.02
Q	10.49	(3.46)	8.12	(2.69)	7.49	0.0001
assignment 1-3 maximum score: 6						
Q_1	3.94	(1.36)	3.26	(1.20)	5.15	0.0001
Q_2	3.61	(1.42)	2.99	(1.25)	4.49	0.0001
Q_3	2.94	(1.69)	1.86	(1.35)	6.93	0.0001

Table C_2 : Group comparison of average scores - posttest

Score Group	5-10	15-20	25-30	35-40	45-50	55-60	65-70	75-80	85-90	95-100
Experimental n = 191	1	5	5	30	34	36	29	27	16	8
Control n = 194	1	8	17	24	23	41	39	19	11	11

$\chi^2 = 7.85$ n.s.

Table C_3 : Distribution table of ACH - pretest

Score Group	5-10	15-20	25-30	35-40	45-50	55-60	65-70	75-80	85-90	95-100
Experimental n = 191		4	9	12	23	29	35	29	30	20
Control n = 194		4	17	23	42	28	29	23	20	8

$\chi^2 = 19.82$ $p = 0.001$

Table C_4 : Distribution table of ACH - posttest

Score / Group	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Experimental n =191			1	2	15	30	48	49	38	8
Control n = 194			1	3	13	38	52	53	24	10

$\chi^2 = 4.96$ n.s.

Table C_5 : Distribution table of ATSI - pretest

Score / Group	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Experimental n = 191			1	6	11	35	49	45	32	12
Control n = 194			0	7	17	37	53	51	25	4

$\chi^2 = 8.43$ n.s.

Table C_6 : Distribution table of ATSI - posttest

Score / Group	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Experimental n = 191					12	37	54	51	27	10
Control n = 194				3	6	30	73	49	25	8

$\chi^2 = 8.89$ n.s.

Table C_7 : Distribution table of ATMA - pretest

Score / Group	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Experimental n = 191				3	9	37	55	44	35	8
Control n = 194				2	14	39	71	40	22	6

$\chi^2 = 6.79$ n.s.

Table C_8 : Distribution table of ATMA - posttest

Score Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Experimental n = 191			2	3	5	17	27	34	32	25	28	11	5	2	0			
Control n = 194			1	4	1	23	27	34	36	32	24	9	0	2	1			

$\chi^2 = 11.62$ n.s.

Table C_9 : Distribution table of Q scores - pretest

Score Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Experimental n = 191			4	3	7	14	12	18	14	24	23	16	17	12	8	12	6	1
Control n = 194	1	1	3	9	19	22	27	39	18	11	17	17	5	4	1			

$\chi^2 = 67.19$ $p = 0.0001$

Table C_{10} : Distribution table of Q scores - posttest

Score Group	0	1	2	3	4	5	6
Experimental n = 191			9	93	64	25	
Control n = 194	1		6	87	78	22	

$\chi^2 = 3.35$ n.s.

Table C_{11} : Distribution table of Q_1 - pretest

Score Group	0	1	2	3	4	5	6
Experimental n = 191	2	2	28	36	59	34	30
Control n = 194		8	53	47	62	14	10

$\chi^2 = 33.12$ $p = 0.0001$

Table C_{12} : Distribution table of Q_1 - posttest

Score / Group	0	1	2	3	4	5	6
Experimental n = 191	9	8	41	73	45	14	1
Control n = 194	11	4	34	77	48	19	1

$\chi^2 = 3.12$ n.s.

Table C_{13} : Distribution table of Q_2 - pretest

Score / Group	0	1	2	3	4	5	6
Experimental n = 191	7	7	26	44	47	50	10
Control n = 194	10	7	43	74	35	24	1

$\chi^2 = 30.58$ $p = 0.0001$

Table C_{14} : Distribution table of Q_2 - posttest

Score / Group	0	1	2	3	4	5	6
Experimental n = 191	20	15	79	55	17	2	3
Control n = 194	28	13	94	48	9	2	0

$\chi^2 = 8.69$ n.s.

Table C_{15} : Distribution table of Q_3 - pretest

Score / Group	0	1	2	3	4	5	6
Experimental n = 191	20	12	47	47	30	15	20
Control n = 194	41	32	63	34	21	1	2

$\chi^2 = 49.23$ $p = 0.0001$

Table C_{16} : Distribution table of Q_3 - posttest

APPENDIX D

COMPARATIVE STUDY :

POPULATION SHIFT FOR ASSIGNMENTS 1 and 3

Experimental group
n = 191

PRE

P \ A	A_1	A_2	A_3
P_1	5%	46.5%	1%
P_2	2%	32%	0.5%
P_3		13%	

POST

P \ A	A_1	A_2	A_3
P_1	14.5%	14%	13%
P_2	5%	18%	17%
P_3		0.5%	16%

Control group
n = 194

PRE

P \ A	A_1	A_2	A_3
P_1	3.5%	44%	1.5%
P_2	1%	39%	
P_3		11%	

POST

P \ A	A_1	A_2	A_3
P_1	27%	17.5%	13%
P_2	7%	19%	7.5%
P_3			5%

Table D_1 : Quality-of-solution distribution for assignment 1

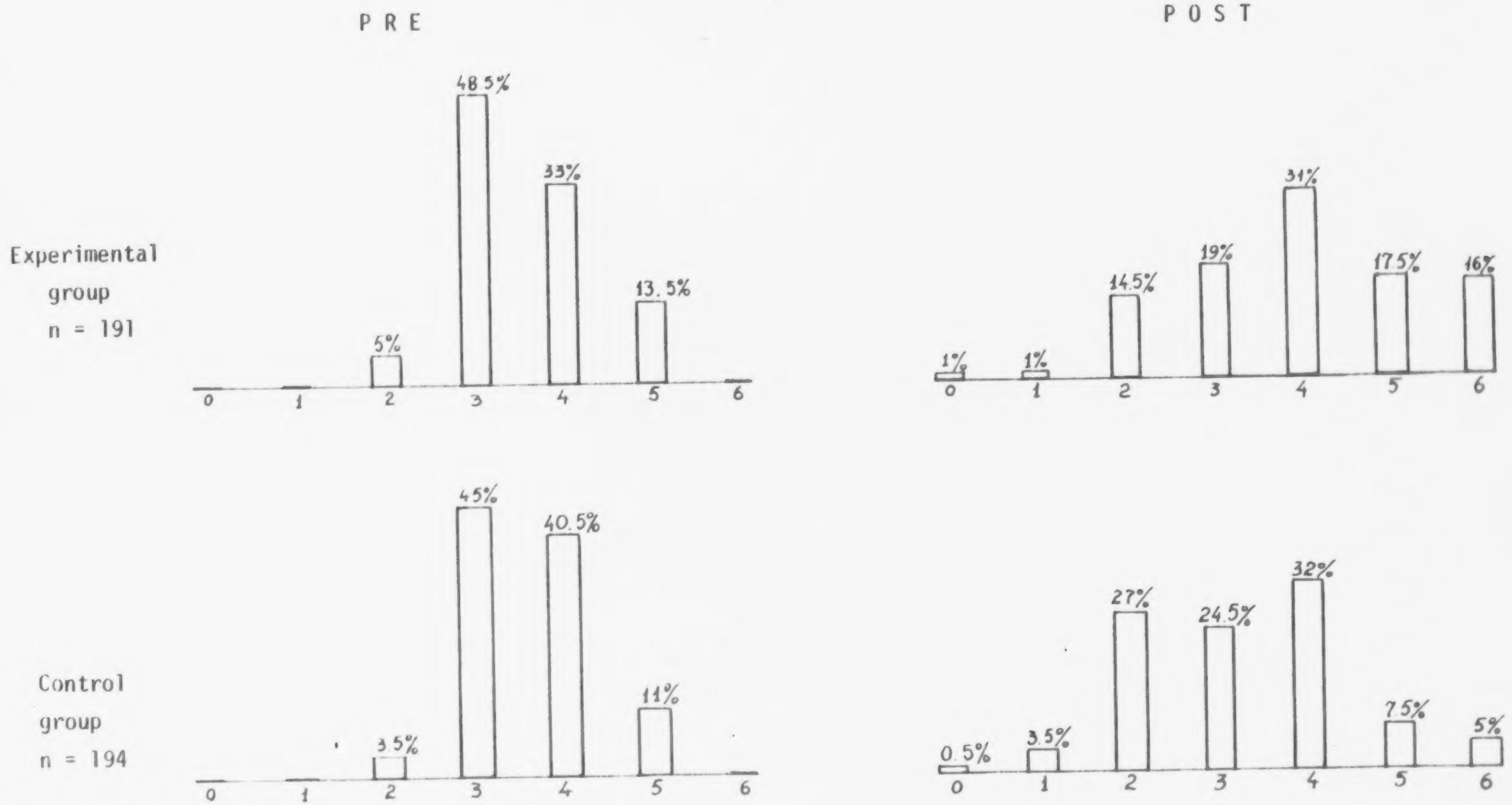
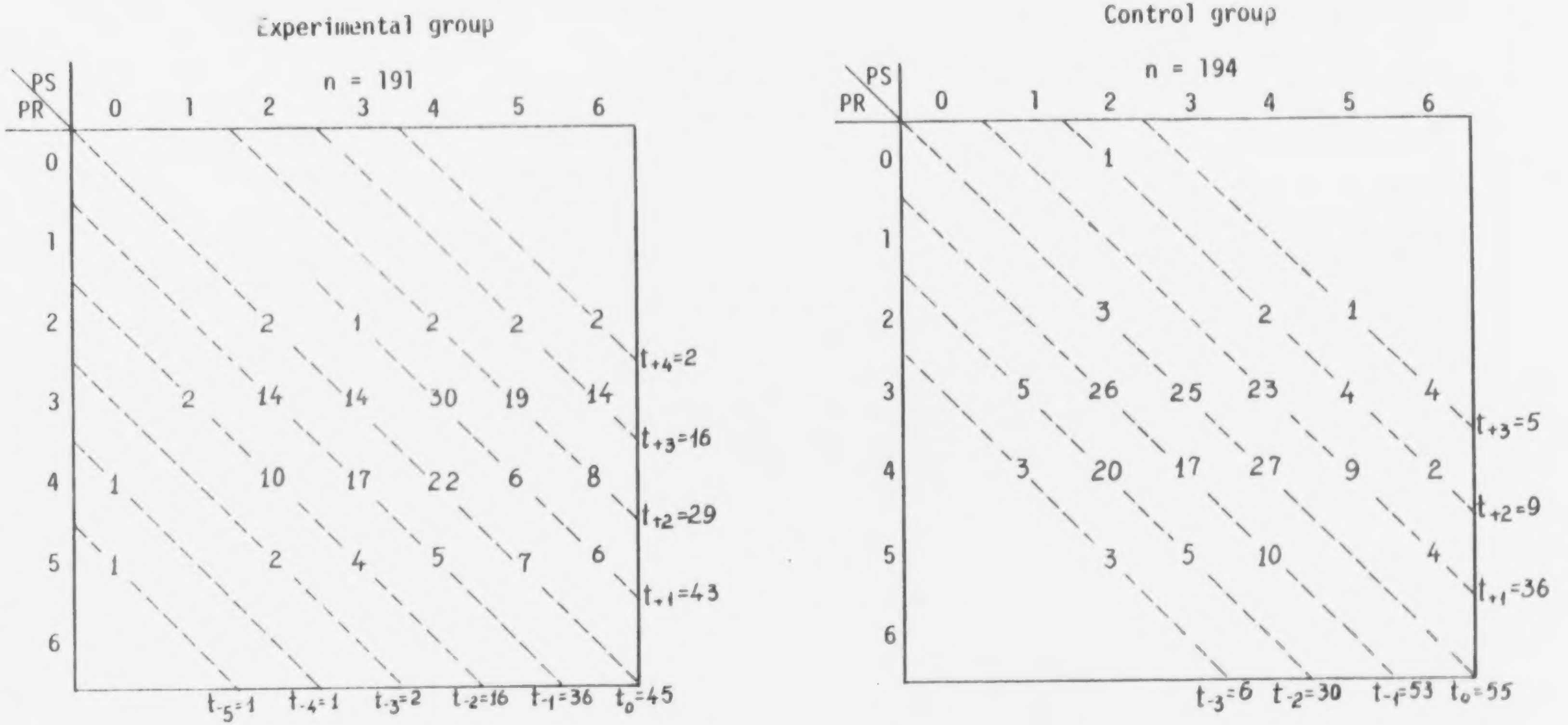


Figure D_2 : Distribution of Q scores for assignment 1



Experimental group

n = 191

Control group

n = 194

PS
PR
0
1
2
3
4
5
6
2
1
2
2
2
2
14
14
30
19
14
1
10
17
22
6
8
1
2
4
5
7
6
$t_{+4}=2$
$t_{+3}=16$
$t_{+2}=29$
$t_{+1}=43$
$t_{-5}=1$
$t_{-4}=1$
$t_{-3}=2$
$t_{-2}=16$
$t_{-1}=36$
$t_0=45$
1
3
2
1
5
26
25
23
4
4
3
20
17
27
9
2
3
5
10
4
$t_{+3}=5$
$t_{+2}=9$
$t_{+1}=36$
$t_{-3}=6$
$t_{-2}=30$
$t_{-1}=53$
$t_0=55$

$S^+ = 90$ $S^- = 56$

T = 6149

E(T) = 8650.5

V(T) = 751.7

U = -3.327

one tailed p = 0.0005

$S^+ = 50$ $S^- = 89$

T = 6021

E(T) = 8687.5

V(T) = 763.5

U = -3.49

one tailed p = 0.0003

Figure D_3 : Population shift for assignment 1

Experimental group
n = 191

PRE

P \ A	A_1	A_2	A_3
P_1	41%	5%	
P_2	24%	9%	0.5%
P_3		0.5%	1.5%

POST

P \ A	A_1	A_2	A_3
P_1	24.5%	10.5%	
P_2	14%	16%	0.5%
P_3		7.5%	10.5%

Control group
n = 194

PRE

P \ A	A_1	A_2	A_3
P_1	48%	1%	
P_2	24%	5%	1%
P_3			

POST

P \ A	A_1	A_2	A_3
P_1	32.5%	4%	
P_2	14%	11%	
P_3		0.5%	1%

Table D_4 : Quality-of-solution distribution for assignment 3

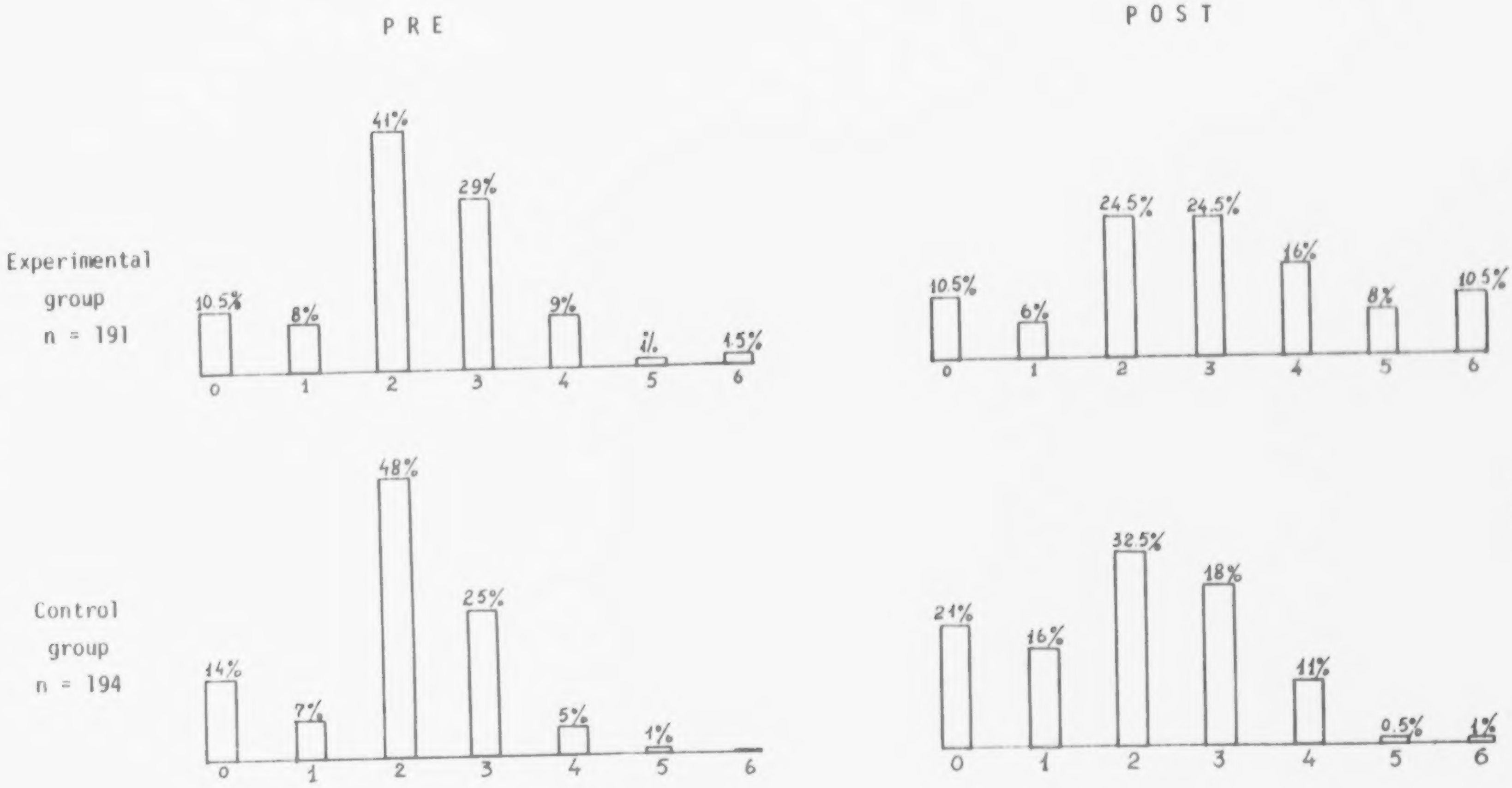
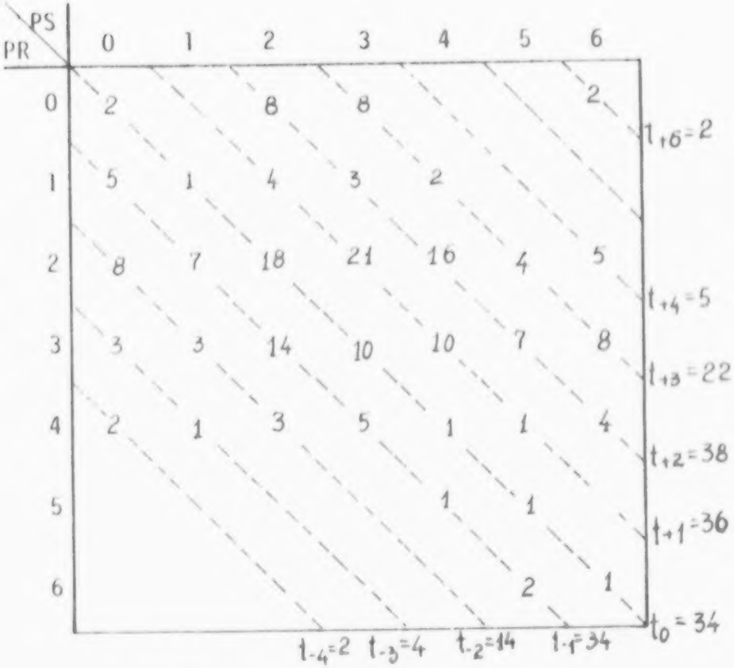


Figure D_5: Distribution of Q scores for assignment 3

Experimental group

n = 191



$t_{+6}=2$
$t_{+4}=5$
$t_{+3}=22$
$t_{+2}=38$
$t_{+1}=36$
$t_0=34$
$t_{-4}=2$
$t_{-3}=4$
$t_{-2}=14$
$t_{-1}=34$

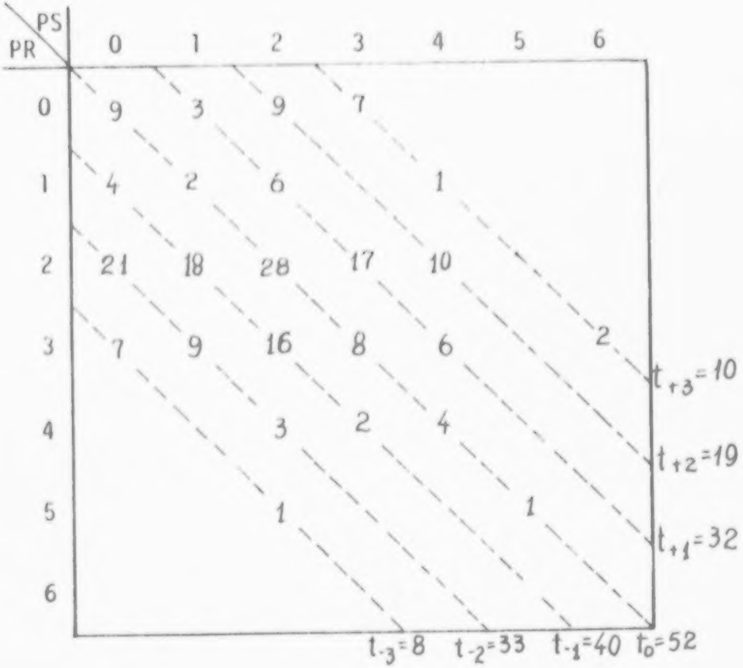
PR \ PS	0	1	2	3	4	5	6
0	2		8	8			2
1	5	1	4	3	2		
2	8	7	18	21	16	4	5
3	3	3	14	10	10	7	8
4	2	1	3	5	1	1	4
5					1	1	
6						2	1

$S^+ = 103$ $S^- = 54$

T = 5240
E(T) = 8870.5
V(T) = 755.8
U = -4.803
one tailed p = 0.00001

Control group

n = 194



$t_{+3}=10$
$t_{+2}=19$
$t_{+1}=32$
$t_{-3}=8$
$t_{-2}=33$
$t_{-1}=40$
$t_0=52$

PR \ PS	0	1	2	3	4	5	6
0	9	3	9	7			
1	4	2	6		1		
2	21	18	28	17	10		
3	7	9	16	8	6		2
4			3	2	4		
5			1			1	
6							

$S^+ = 61$ $S^- = 81$

T = 7546.5
E(T) = 8768.5
V(T) = 768.2
U = -1.590
one tailed p = 0.056

Figure D_6 : Population shift for assignment 3

APPENDIX E

POPULATION SHIFT IN THE COMMUNICATION STUDY

GR1
n = 84

1st

P \ A	A_1	A_2	A_3
P_1	7%	32%	
P_2	1%	40%	
P_3		18%	2%

2nd

P \ A	A_1	A_2	A_3
P_1	10%	17%	8%
P_2	2%	39%	8%
P_3		3%	12%

GR2
n = 83

1st

P \ A	A_1	A_2	A_3
P_1	16%	23%	5%
P_2	6%	30%	6%
P_3		3%	10%

2nd

P \ A	A_1	A_2	A_3
P_1	1%	30%	
P_2	2%	35%	5%
P_3		17%	10%

Table E_1 : Quality-of-solution distribution for assignment 1

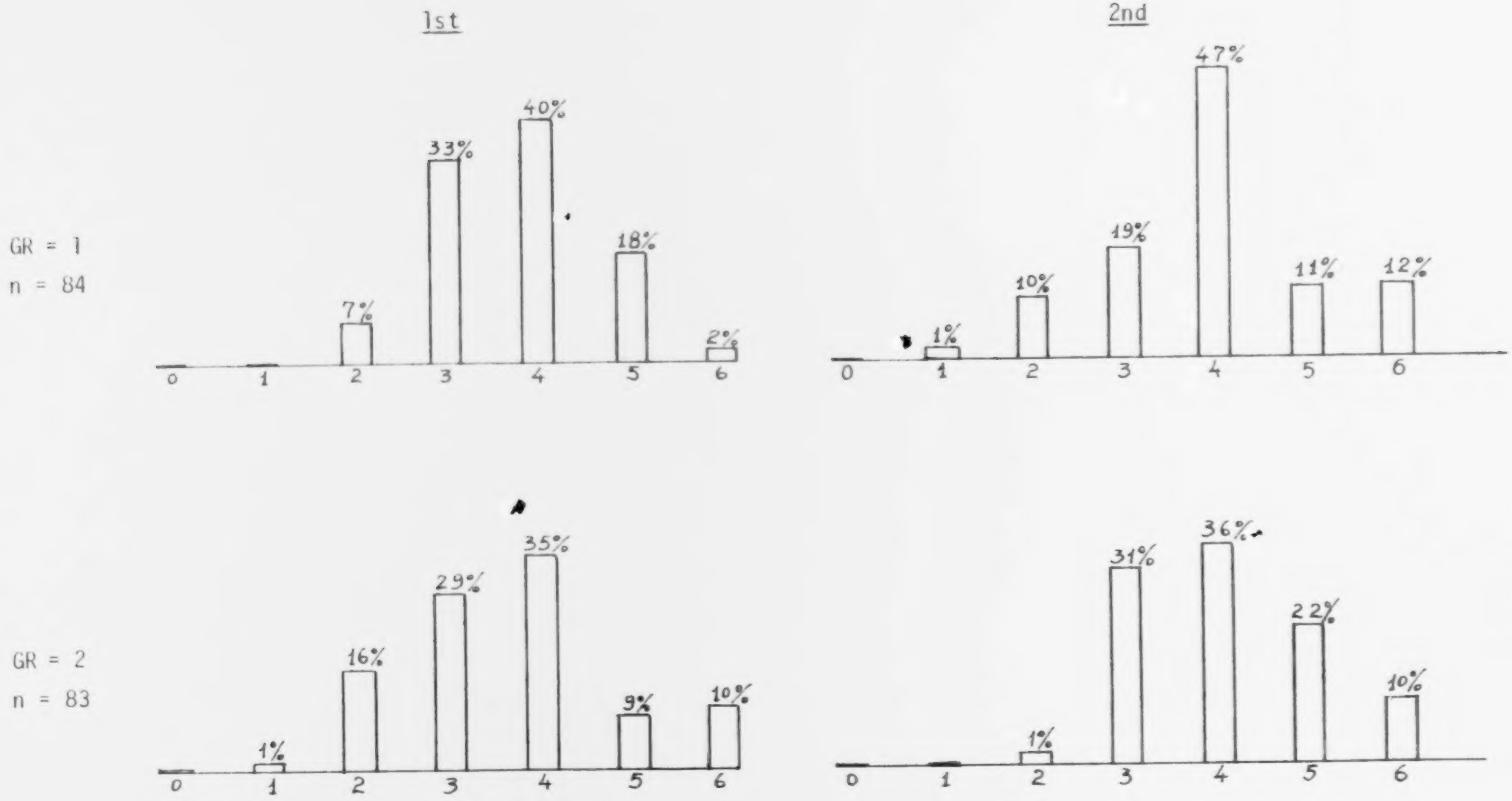
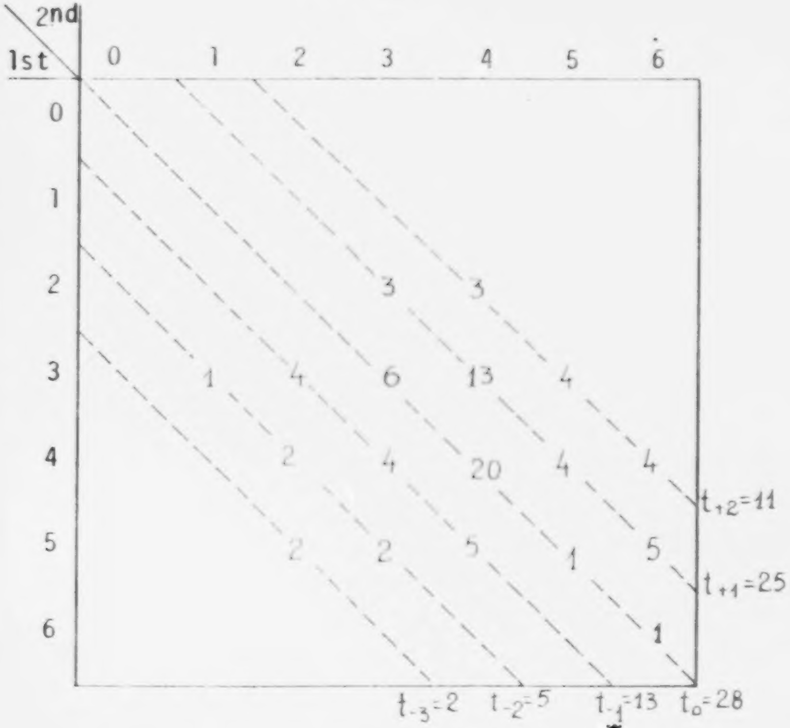


Figure E_2 : Distribution of Q scores for assignment 1

GR1

n = 84



2nd
1st
$t_{+2}=11$
$t_{+1}=25$
$t_{-3}=2$
$t_{-2}=5$
$t_{-1}=13$
$t_0=28$

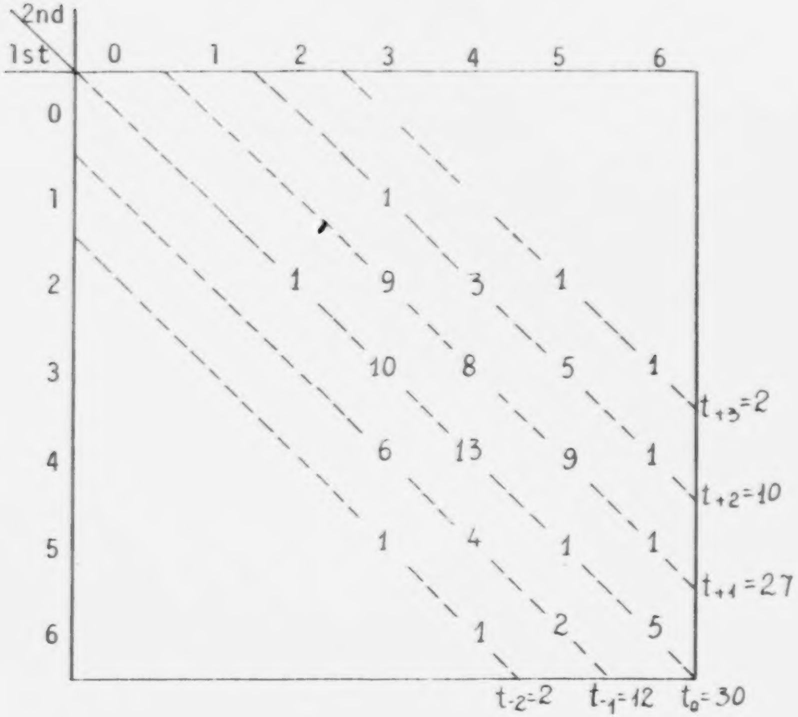
1st \ 2nd	0	1	2	3	4	5	6
0							
1							
2				3	3		
3		1	4	6	13	4	
4			2	4	20	4	4
5			2	2	5	1	5
6							1

$S^+ = 36$ $S^- = 20$

T = 1157
E(T)= 1582.5
v(T)= 217.2
U = -1.957
one tailed p = 0.025

GR2

n = 83



2nd
1st
$t_{+3}=2$
$t_{+2}=10$
$t_{+1}=27$
$t_{-2}=2$
$t_{-1}=12$
$t_0=30$

1st \ 2nd	0	1	2	3	4	5	6
0							
1				1			
2			1	9	3	1	
3				10	8	5	1
4				6	13	9	1
5				1	4	1	1
6					1	2	5

$S^+ = 39$ $S^- = 14$

T = 751
E(T)= 1510.5
v(T) = 218.1
U =-3.487
one tailed p = 0.0001

Figure E_3 : Population shift for assignment 1

1st 2nd

GR1
n = 84

P \ A	A_1	A_2	A_3
P_1	19%	36%	1%
P_2	6%	27%	
P_3		3%	1%

P \ A	A_1	A_2	A_3
P_1	10%	7%	4%
P_2	7%	32%	13%
P_3		8%	12%

GR2
n = 83

P \ A	A_1	A_2	A_3
P_1	19%	16%	
P_2	10%	31%	4%
P_3		1%	4%

P \ A	A_1	A_2	A_3
P_1	11%	21%	7%
P_2	1%	23%	12%
P_3		11%	7%

Table E_4 : Quality-of-solution distribution for assignment 2

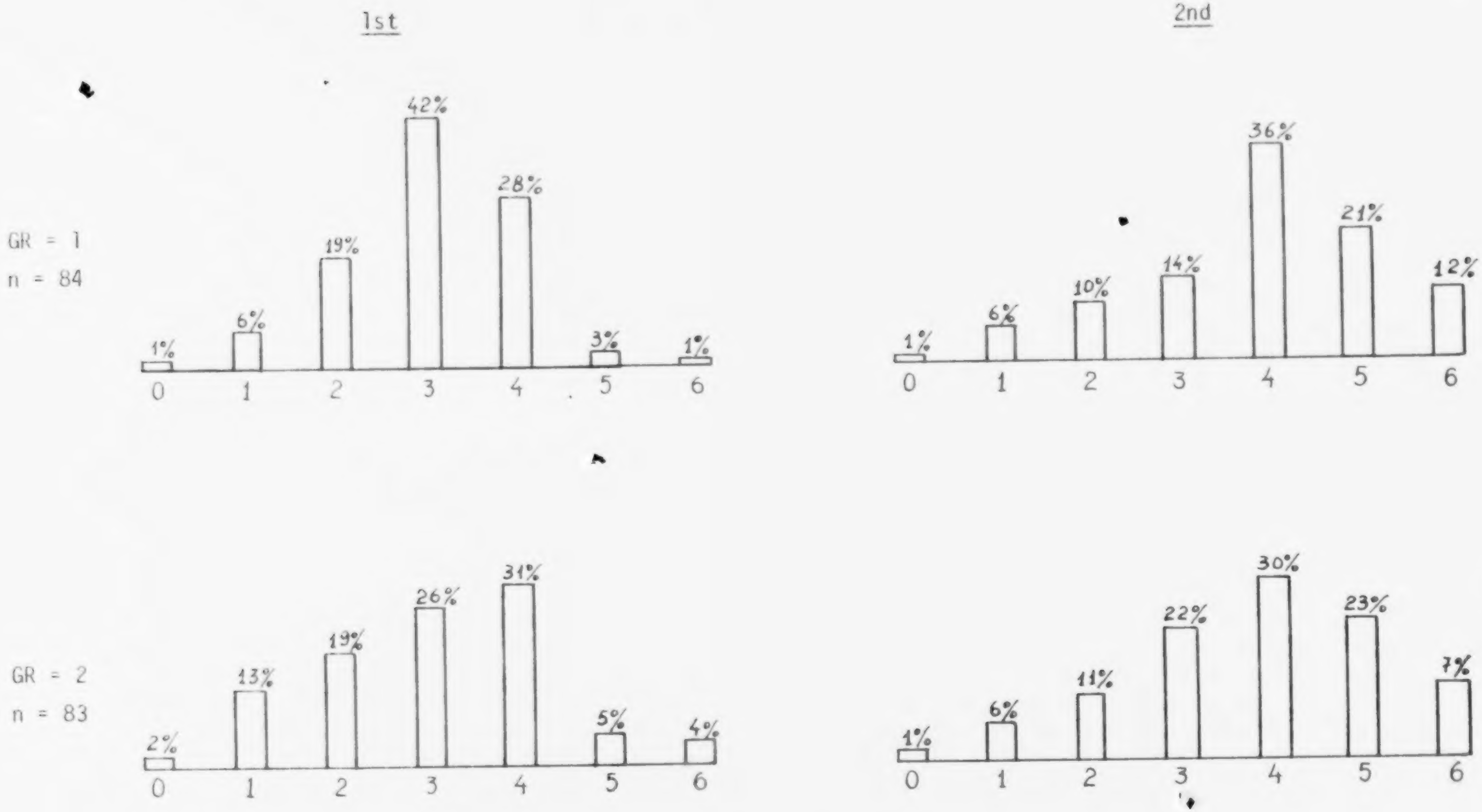


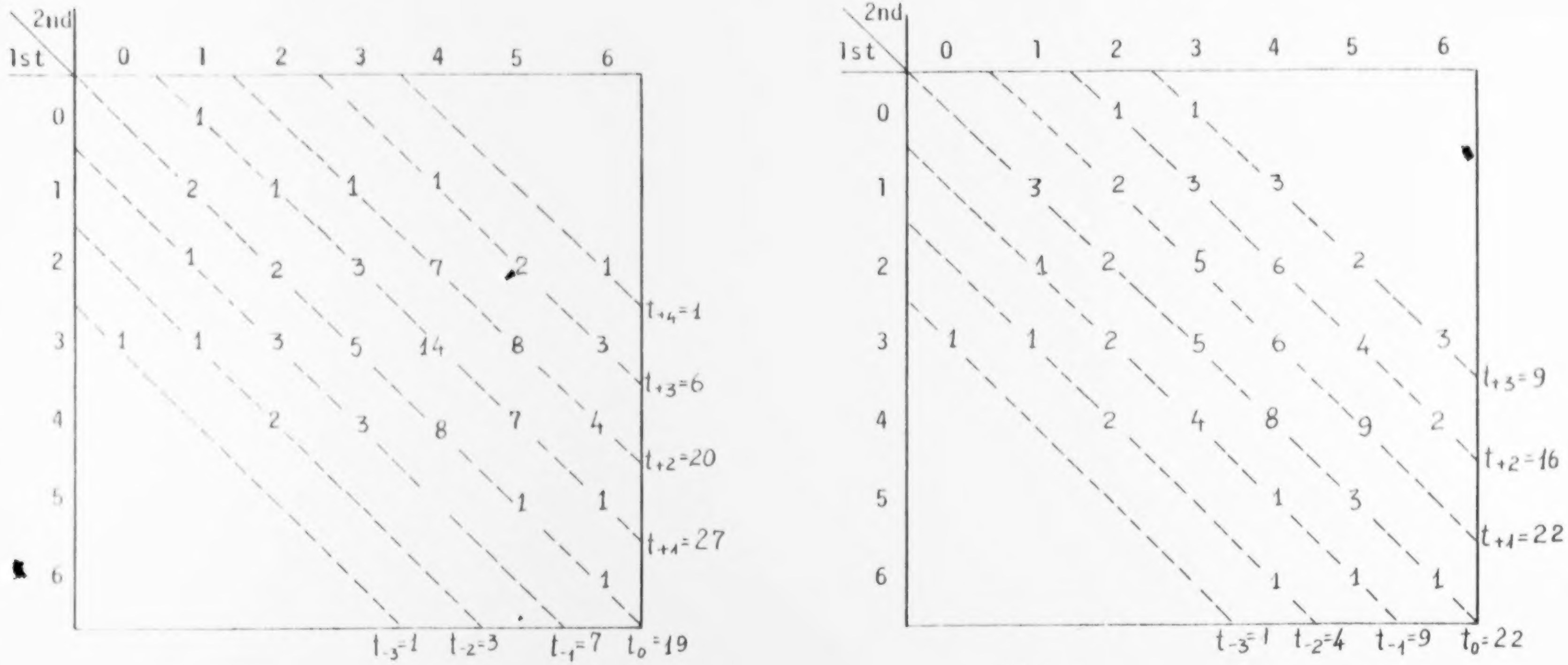
Figure E_5 : Distribution of Q scores for assignment 2

GR1

n = 84

GR2

n = 83


2nd
1st
$t_{+4}=1$
$t_{+3}=6$
$t_{+2}=20$
$t_{+1}=27$
$t_{-3}=1$
$t_{-2}=3$
$t_{-1}=7$
$t_0=19$
$t_{+3}=9$
$t_{+2}=16$
$t_{+1}=22$
$t_{-3}=1$
$t_{-2}=4$
$t_{-1}=9$
$t_0=22$

GR1

1st \ 2nd	0	1	2	3	4	5	6
0		1					
1		2	1	1	1		
2		1	2	3	7	2	1
3	1	1	3	5	14	8	3
4			2	3	8	7	4
5						1	1
6							1

GR2

1st \ 2nd	0	1	2	3	4	5	6
0			1	1			
1		3	2	3	3		
2		1	2	5	6	2	
3	1	1	2	5	6	4	3
4			2	4	8	9	2
5					1	3	
6					1	1	1

$S^+ = 54$ $S^- = 11$

T = 537.5
E(T)= 1690
v(T)= 220
U = -5.22
one tailed p = 0.00001

$S^+ = 47$ $S^- = 14$

T = 594.5
E(T)= 1616.5
v(T)= 216.3
U =-4.726
one tailed p = 0.00001

Figure E_6 : Population shift for assignment 2

GR1
n = 84

1st

P \ A	A_1	A_2	A_3
P_1	50%	5%	1%
P_2	14%	8%	
P_3			1%

2nd

P \ A	A_1	A_2	A_3
P_1	40%	4%	
P_2	17%	15%	
P_3		2%	4%

GR2
n = 83

1st

P \ A	A_1	A_2	A_3
P_1	43%	11%	
P_2	8%	12%	
P_3			

2nd

P \ A	A_1	A_2	A_3
P_1	40%	7%	
P_2	22%	11%	6%
P_3			5%

Table E_7 : Quality-of-solution distribution for assignment 3

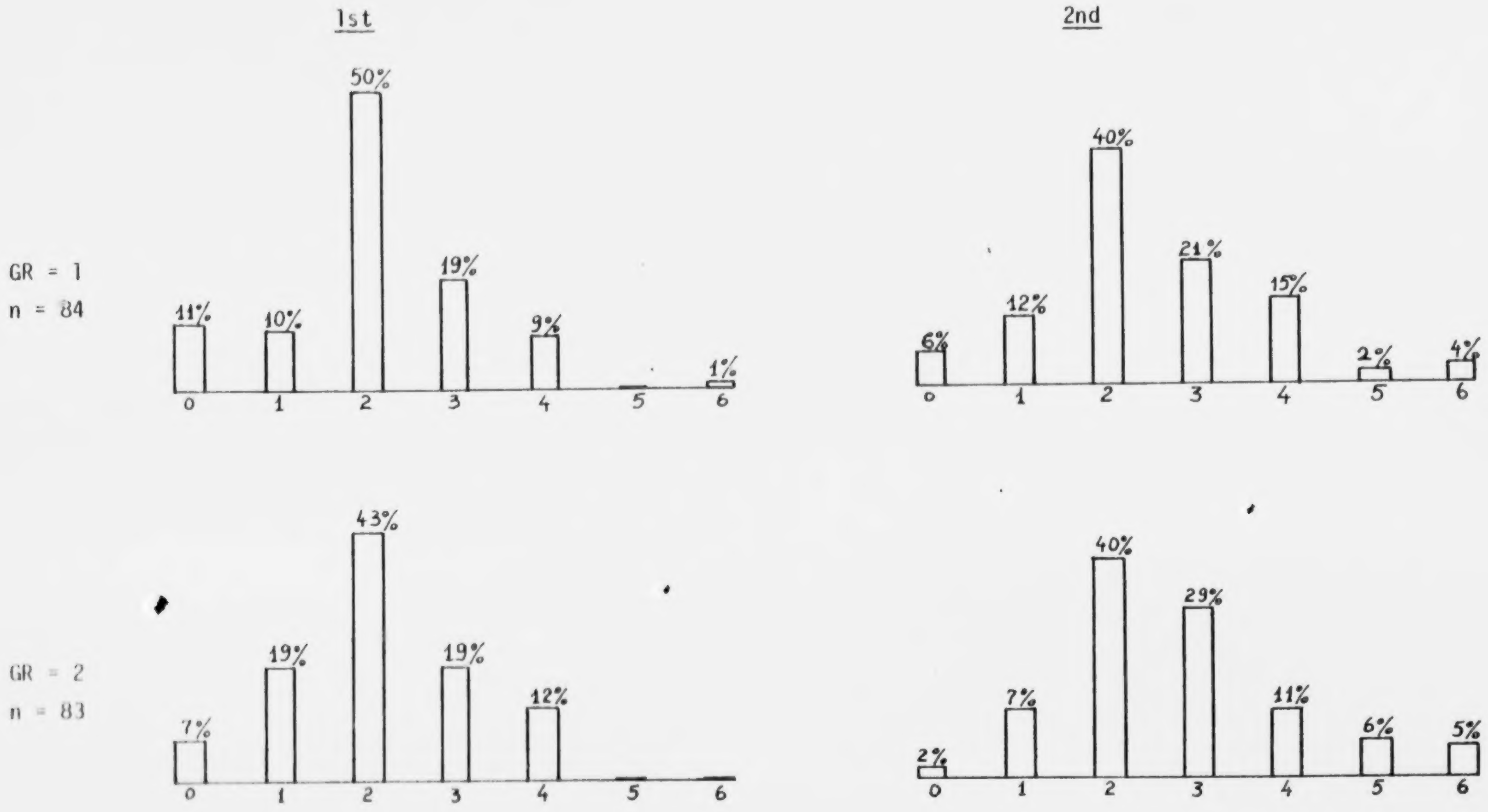


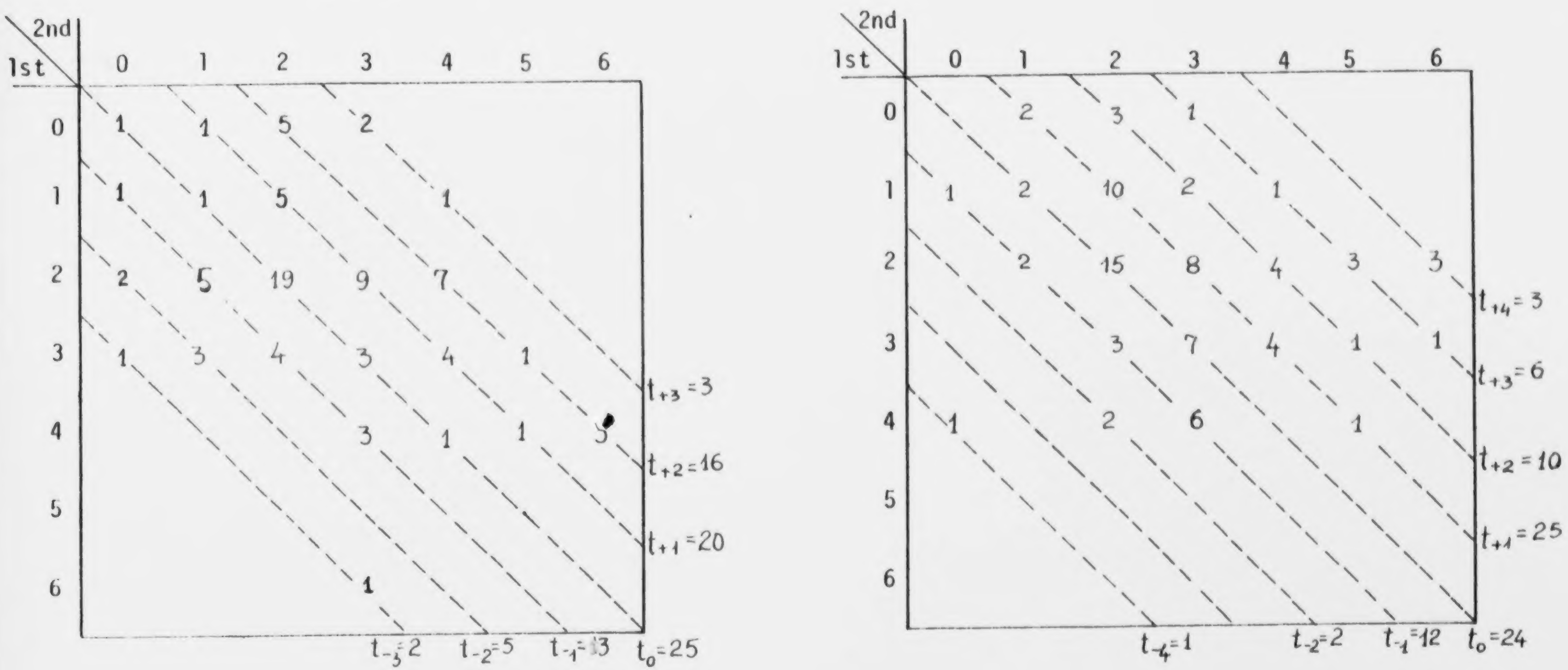
Figure E_8 : Distribution of Q scores for assignment 3

GR1

n = 84

GR2

n = 83



GR1 (n = 84)

1st \ 2nd	0	1	2	3	4	5	6
0	1	1	5	2			
1	1	1	5		1		
2	2	5	19	9	7		
3	1	3	4	3	4	1	
4				3	1	1	3
5							
6				1			

$t_{+3} = 3$, $t_{+2} = 16$, $t_{+1} = 20$, $t_0 = 25$, $t_{-1} = 13$, $t_{-2} = 5$, $t_{-3} = 2$

GR2 (n = 83)

1st \ 2nd	0	1	2	3	4	5	6
0		2	3	1			
1	1	2	10	2	1		
2		2	15	8	4	3	3
3			3	7	4	1	1
4	1		2	6		1	
5							
6							

$t_{+4} = 3$, $t_{+3} = 6$, $t_{+2} = 10$, $t_{+1} = 25$, $t_0 = 24$, $t_{-1} = 12$, $t_{-2} = 2$, $t_{-4} = 1$

$S^+ = 39$ $S^- = 20$

T = 1055

E(T) = 1622.5

v(T) = 218.9

U = 2.591

one tailed p = 0.0048

$S^+ = 44$ $S^- = 15$

T = 732

E(T) = 1593

v(T) = 214.9

U = -4.006

one tailed p = 0.00003

Figure E_9 : Population shift for assignment 3

APPENDIX F

A MODEL FOR COMPARING TEACHER EXPECTATION AND STUDENT PERFORMANCE

Appeared in :

Journal for Research in Mathematics Education,
Vol. 12, No. 2, March 1981, 142-151.

A method of analyzing tests using the teacher's predictions

by

Nurit Zehavi and Maxim Bruckheimer

Science Teaching Center

Weizmann Institute

Rehovot, Israel

Abstract

As a part of the evaluation of the mathematics curriculum for junior high school, a test was given to 28 ninth grade classes. Each teacher was asked to classify each item on a three point cognitive scale, and to predict the success of the class on each item on a four point scale. In addition to the usual methods of test analysis, a three variable graphical representation was drawn for each class and for each question. The three variables presented were the students' success, teacher-assigned cognitive level, and success as predicted by the teacher. The graphic representation connects the three constituents of student, teacher and program.

The teacher can be involved in an evaluation procedure if account is taken of the teacher's judgement of cognitive level and prediction of difficulty of the test items. This data can be correlated with the students' performance to reveal diagnostic information not usually available by "classical" methods.

The method we shall describe was applied to a mathematics test given to ninth grade students. The three variables, the students' success, the teacher-assigned cognitive level and success as predicted by the class teacher were presented graphically. The graphic representation gives information about aspects of the teacher-student interaction for each class and the test population as a whole. The results also contribute to the evaluation of the program and provide useful information for the developers, the implementors, and the teachers.

Background

In the implementation of new curricula the relationship between the curriculum developers and the teachers in the classroom is of great importance. The developers have certain intentions and objectives which influence choice of subject matter, media and recommended teaching strategies. They try to convey their intentions to the teachers in the process of implementing the curriculum. At the same time, the teacher has background, education, experience, and opinion that determine an individual teaching strategy. The interaction of all these variables with the specific class is the actual process that the evaluator wishes to study and analyse in order to improve the process and product of students' learning. Many of the concepts and parameters used in the evaluation are essentially not objective, but subject to the influence of these variables. This is the basic premise of the present study which incorporates the subjective judgement of the teacher.

The Method

The Bloom taxonomy was adapted to mathematics by Avital (Avital and Shettleworth, 1968). The cognitive domain was classified into a simplified three level form (defined in Avital)

(i) technique

(ii) comprehension and application (algorithmic thinking)

(iii) analysis and synthesis.

What we call "technique" is included in Avital's category of knowledge.

As part of our in-service teacher training program we ran a series of workshops on questionnaire and test preparation. We offered the teachers collections of test-items and asked them to classify them according to content and cognitive level and to use them to prepare a test appropriate to their classes. We explained the cognitive levels following Avital. It was very clear that the teachers disagreed on many questions, in spite of the fact that all of them teach exactly the same program and use the same materials. This strengthened the view that we should involve the teachers, by our requesting them to judge the cognitive level of the questions in relation to their classes, and our incorporating this information into the subsequent analysis.

A further parameter used in the analysis of tests is the difficulty index of each question. This also depends on the teacher and the class, hence it seems reasonable to assume that we can obtain valuable information by asking the teacher to predict the success rate on each question for his class. A teacher "in touch", with the students in his class, should be able to predict "roughly" how many will succeed on a specific item. For the teacher's prediction we suggested a four point scale.

1 - less than 25% will answer correctly;

2 - 25% - 50% will answer correctly;

3 - 50% - 75% will answer correctly;

4 - more than 75% will answer correctly.

The Test

As a part of the evaluation of the mathematics curriculum developed by the Department of Science Teaching of the Weizmann Institute, an 18-item test was given to 28 ninth grade classes (n=670). The test included questions on material studied in the three years of junior high school and covered the following topics:

(a) functions;

(b) open phrases;

(c) sets.

Most of the items were chosen from an item bank which had been constructed from tests used in previous years. For each item we had a difficulty and discrimination index. The composition of the test reflected the balance in the Grade 9 curriculum between classical and "new math" content, as well as a spread of items of different cognitive levels as judged by experts.

The mean score on the test as applied to our sample was 55, with standard deviation of 20, and Kuder-Richardson reliability index 0.8.

As expected there was considerable variation in the teachers' estimation of both cognitive level and difficulty. It is worth mentioning that most of the teachers were experienced, and had taught the particular class in all three years of junior high school. They had no previous experience in this form of estimation, but they had all previous experience of the cognitive taxonomy and classification of questions accordingly.

The data were presented graphically as follows:

(i) the x-axis was used to represent success on a particular question as predicted by the teacher;

(ii) the y-axis was used to represent the actual success rate on a question;

(iii) the cognitive levels (as estimated by the class teacher) were represented by colors or by the following shapes:

technique - △
comprehension and application - □
analysis and synthesis - ⬡

Discussion

The subjective judgements of the teachers increase the information obtained from the test. An analysis was done from the point of view of the evaluator of the program and its implementation. In this case the evaluators were also the developers and implementors.

To illustrate the sort of information that can be obtained from such representations, two types of analysis will follow:

A. of one question over all the classes.

B. of one class over all the questions in the test.

Analysis of individual question over all the classes

The procedure will be illustrated by the following three items, the associated graphs, and the discussion.

Example 1: Question 5, from the section of the test on functions.

Consider the family of linear functions $y = 2x + b$, where b can be any real number.

Which of the following sentences is true?

Among the graphs of the functions:

a) there is a pair which meet on the x-axis.

b) there is a pair which meet on the y-axis.

c) there is a graph parallel to the x-axis.

d) there is a graph which passes through the origin.

Looking at Figure 1 one can see that the question was not very difficult in any of the classes and only in one class was it very easy. The mean for the whole population was 54. Most of the teachers judged it to be at the comprehension level and predicted it to be of medium difficulty, with a significant minority who ranked it very easy or very difficult. Teachers who predicted it to be very difficult also tended to classify it at the level of analysis, but their students did considerably better than their prediction.

In general the "points" (especially those at the comprehension level) are not far from the line $y = x$ which implies that the teacher prediction on this question is close to reality.

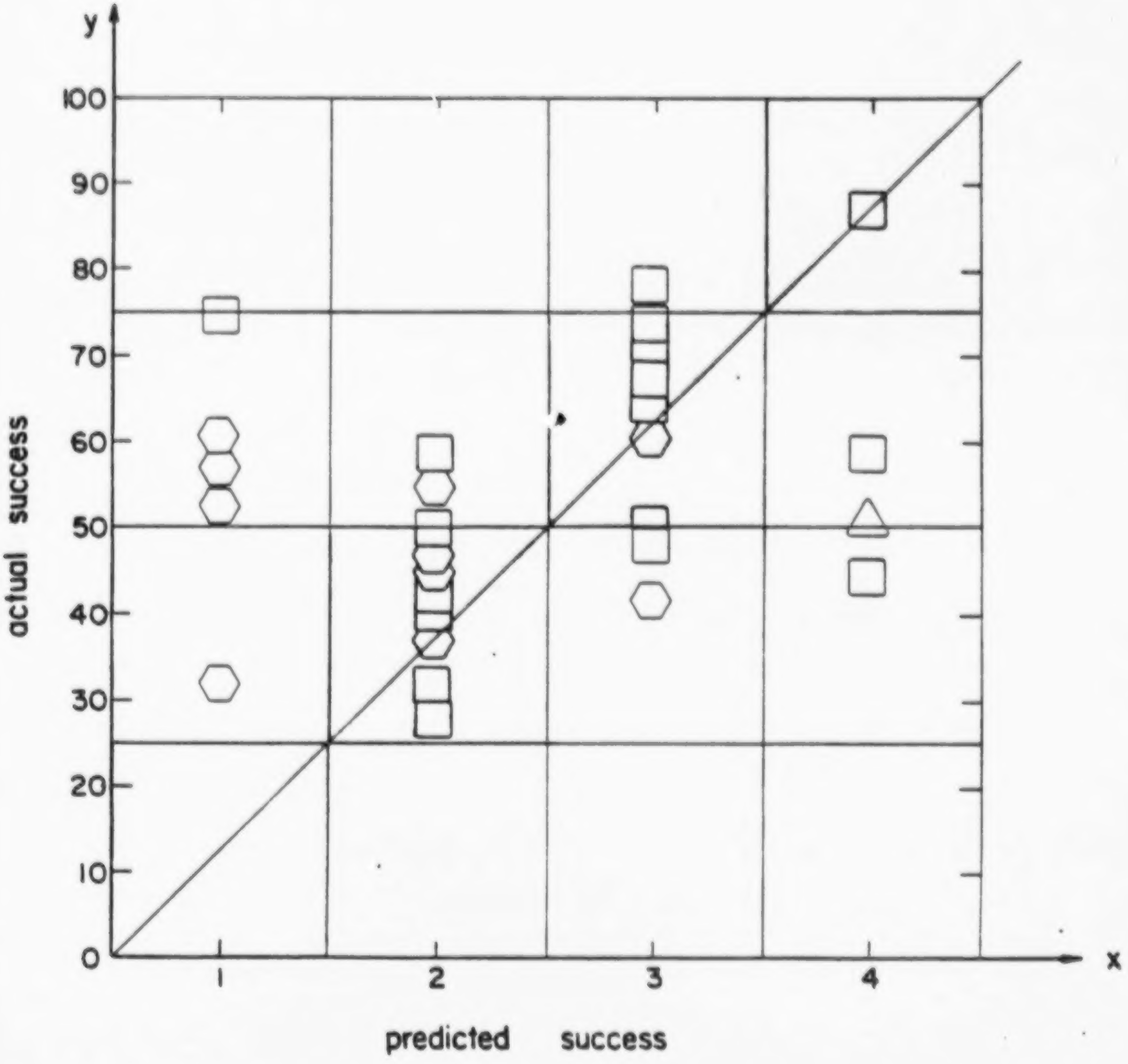


Fig. 1. Graphical representation of data for question 5.

The relation between the parameters of a linear function and its graph is a subject dealt with extensively in the program. However, the phrase "family of linear functions" does not appear. This is probably the reason why nearly all the teachers classified it at the level of comprehension or above. The results show, nevertheless, that the students had mastered the material sufficiently for them to answer the question successfully in spite of its somewhat unfamiliar and more mathematical form. In this particular example, there were only a few cases of serious over- or underestimation (represented by the shapes outside the squares on the diagonal). The following two examples, however, show a quite different picture.

Example 2: Question 7, from the section on open phrases.

The expression (open phrase)

$(a + b)^2 - (a - b)^2$

is equivalent to

a) $2ab$ b) $2b^2$ c) $4ab + 2b^2$ d) $4ab$

This is obviously a technique question according to most of the teachers. However, something is quite clearly wrong here: the mean score for the whole population was 40. This program, like many other "modern" programs has been under attack for neglecting techniques. Maybe for algebraic expressions there is still something to be done. But, what is interesting from the graph (Fig.2) is that the teachers do not seem to be aware of the difficulty and, generally, they expect their students to do well. Maybe the teachers remember this topic as a common technique from their own studies and hence, by wishful thinking, transfer their knowledge to their students. Clearly here is something to be checked and the general direction of the investigation is also clear.

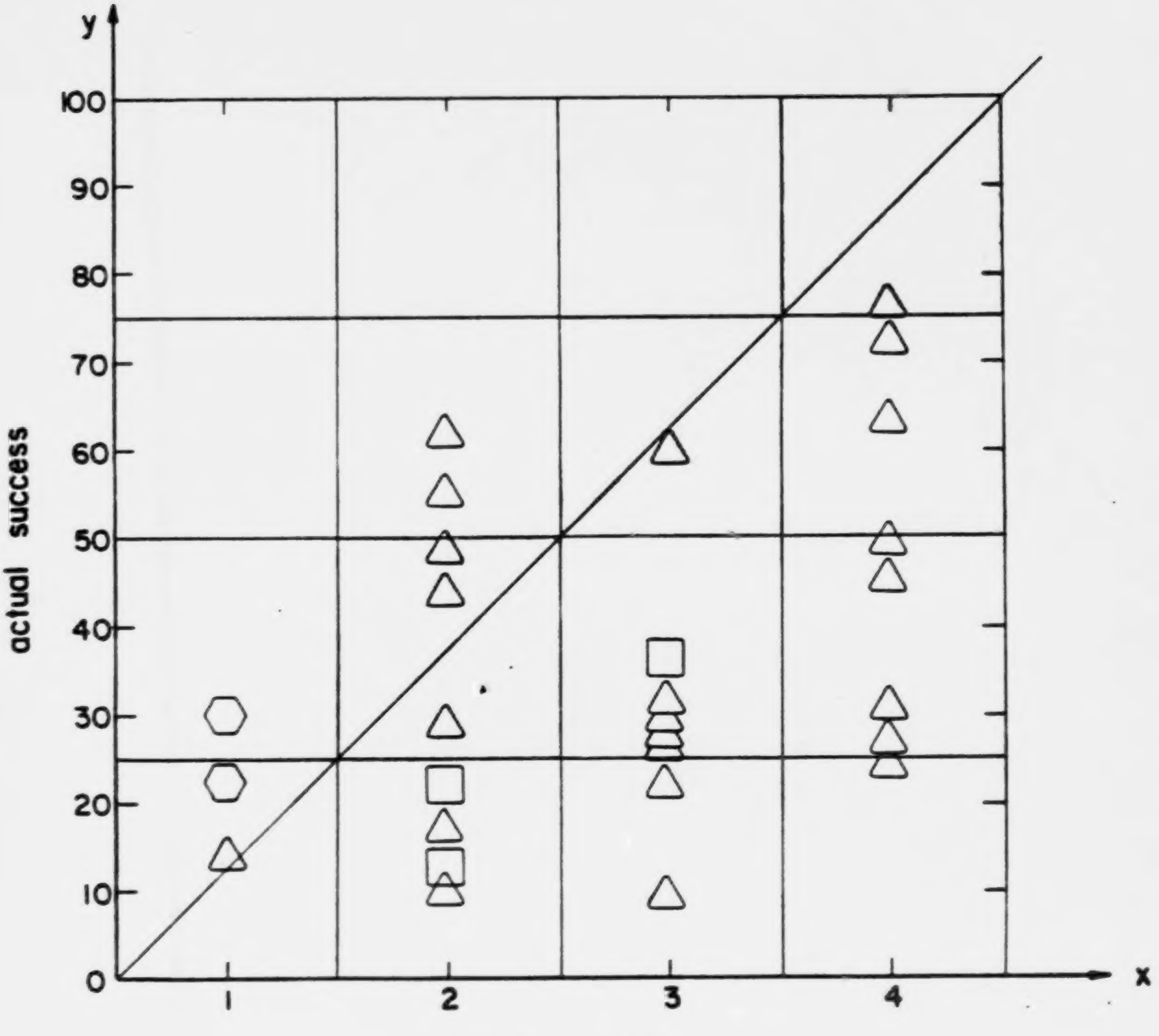


Fig. 2. Graphical representation of data for question 7.

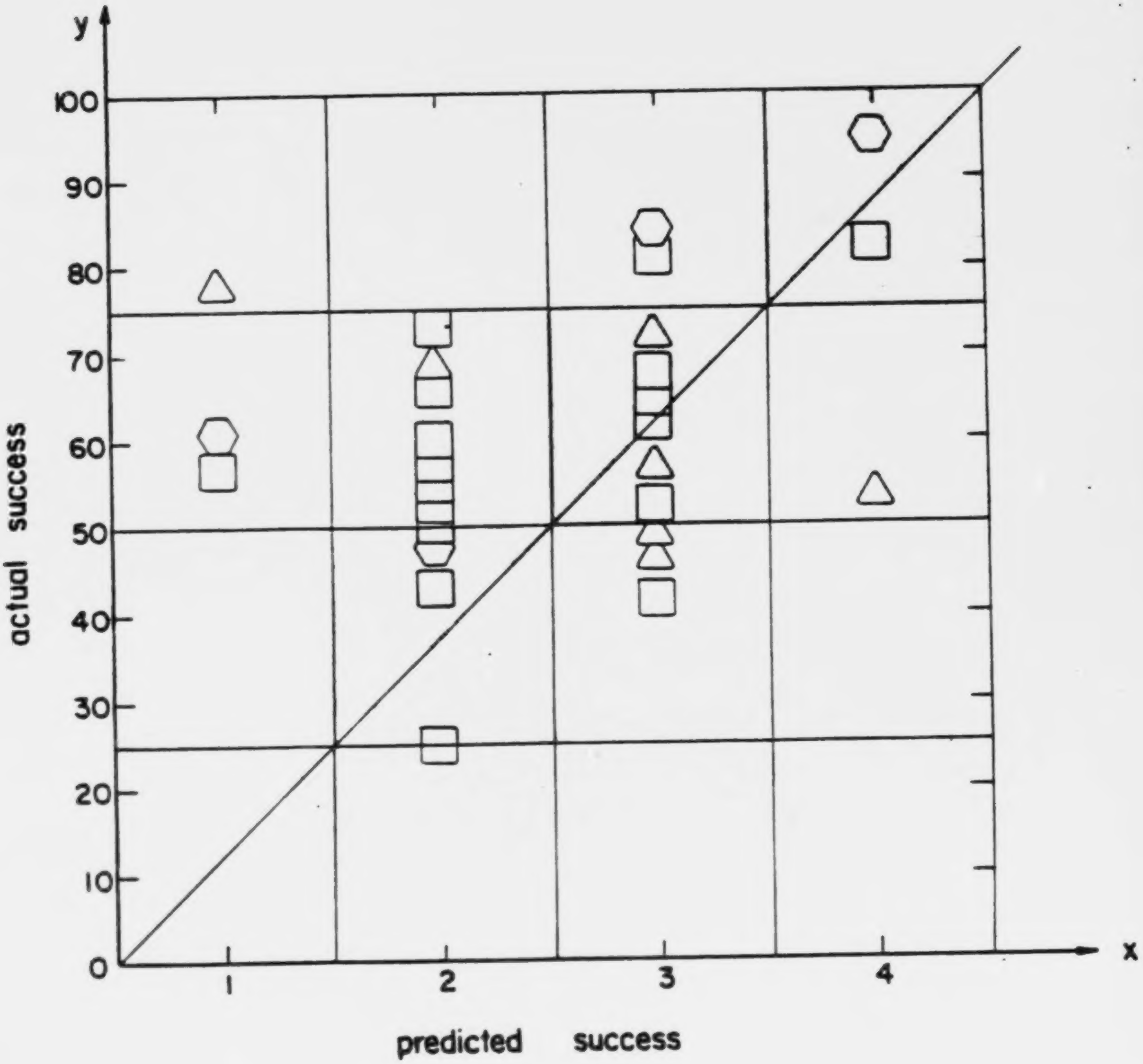


Fig. 3. Graphical representation of data for question 13.

Example 3: Question 13, from the section on open sentences and truth sets.

The truth set of the open sentence (inequality) $|x| + x < 0$ *is:*

a) $\{\ \}$ b) $\{0\}$ c) $\{x|x > 0\}$ d) $\{x|x < 0\}$

The majority of teachers classified the question as comprehension, but overestimated the difficulty (mean score 65). This might seem to suggest that the material is treated adequately in the program, but the teachers have some "mental block" with regard to it. (In fact, comments of the form, "when I was at school we didn't do sets or inequalities" were heard from teachers.) This suggests that further investigation and possible remedial action should be centered on the teacher.

Analysis of one class over all the questions

So far the discussion has exemplified the sort of contribution this form of analysis can make to content evaluation and implementation. In this section, we shall exemplify what can be learnt and done with regard to the teacher-student interaction.

In general, the teachers in this study predicted their students' success on most test items with a reasonable degree of accuracy. For example, consider the following class (a) (mean score 66).

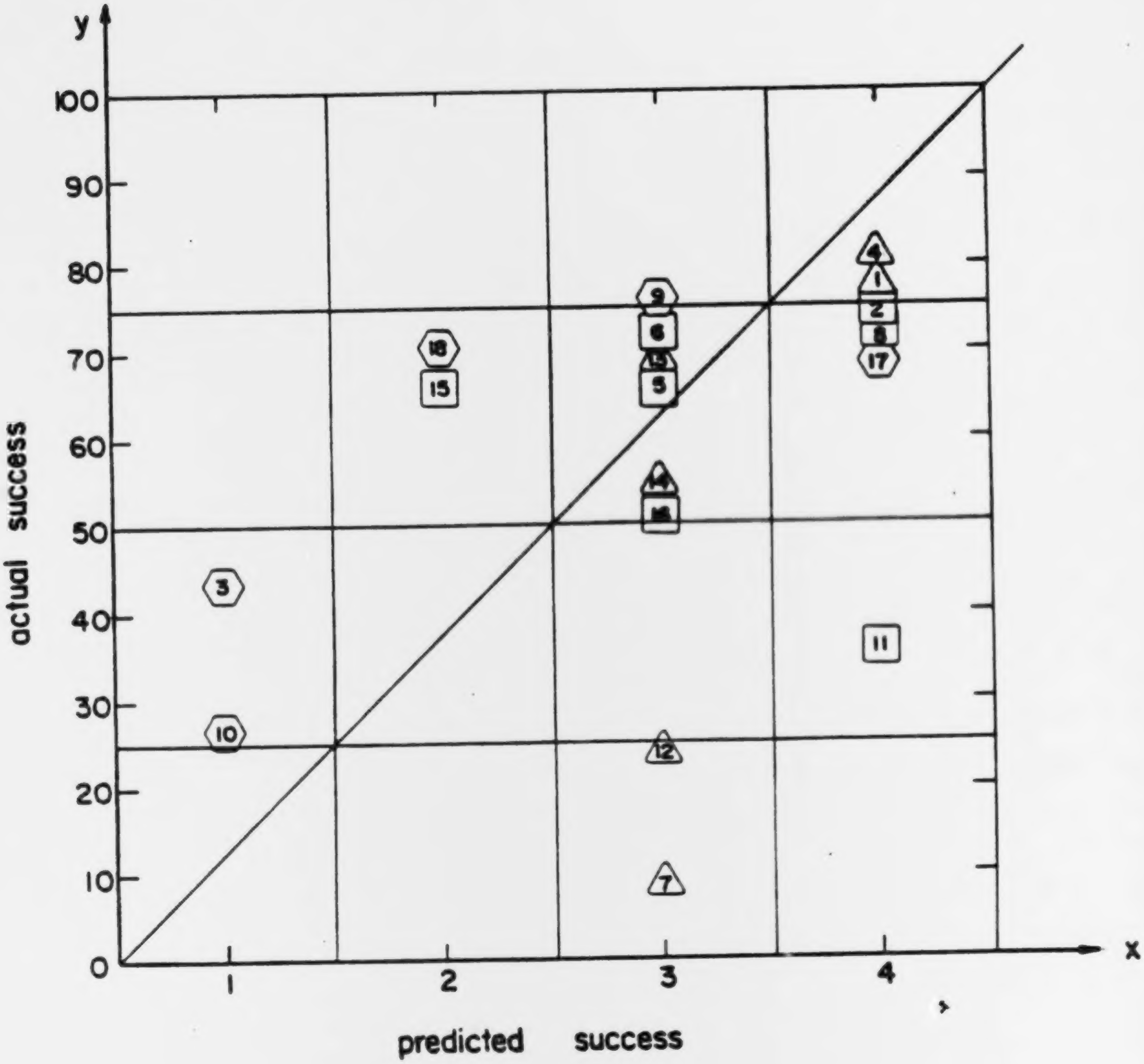


Fig. 4. Graphical representation of data for class (a)

Note: In this, and the following two figures, the actual success is the mean score of the specific class for each item in the test. The number of the items, 1-18, appear within the appropriate shapes.

In this graph, there is a clear relation between the estimated and the actual success. The teacher judged 5 questions to be at the highest cognitive level but expected some of them not to be very difficult for his students and he was correct. The technique and comprehension questions were, on the whole, estimated to be easier, but technique not easier than comprehension. From all the above one gets the impression that class (a) is a "good" one, that the teacher is experienced and is "in touch" both with his class and the program. But note, that there are items, for example item 7 (see discussion above), to which it would be worth drawing the teacher's attention.

The following two examples, illustrate somewhat exceptional situations.

In class (b) one can see immediately that there is a problem (Fig.5).There is a clear underestimation of the students' ability. Except for one single technique question, the test was classified about fifty-fifty at the comprehension and synthesis levels. Nevertheless, the class is clearly a "good" one, and obtained a mean score (63), almost the same as class (a). For some reason or other, the teacher does not seem to be in touch with his class. Speculation as to the reason can be very rewarding, but in the end observation in the classroom is essential. In this particular case the teacher, although experienced, generally taught from the blackboard. He was sure that the best way to be certain that the students learn was to explain and explain again.

He did not believe that students were capable of doing anything without his first explaining every detail.

Class (c) is another clear example of a class and a teacher out of touch with each other (Fig.6).

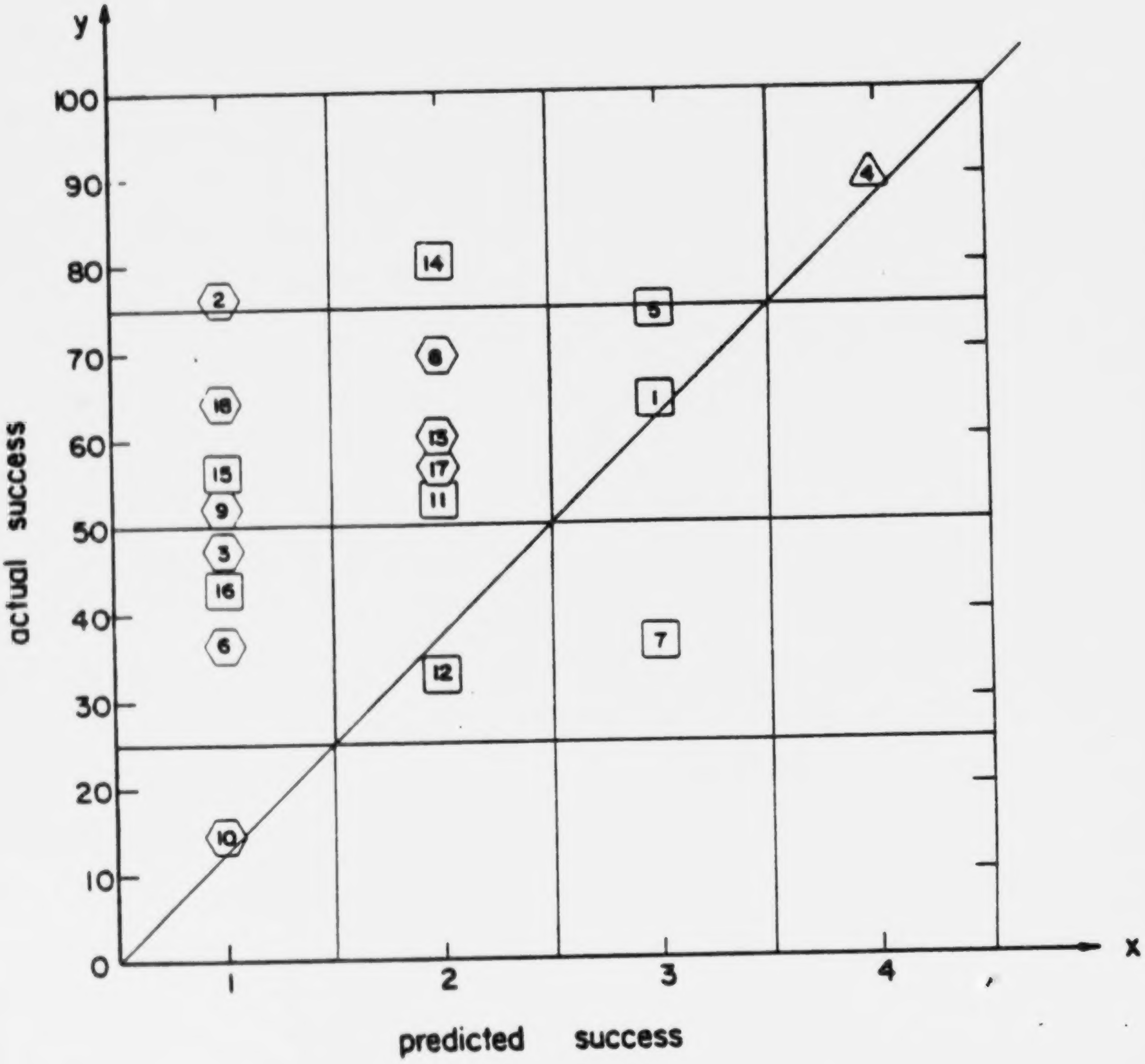


Fig. 5. Graphical representation of data for class (b)

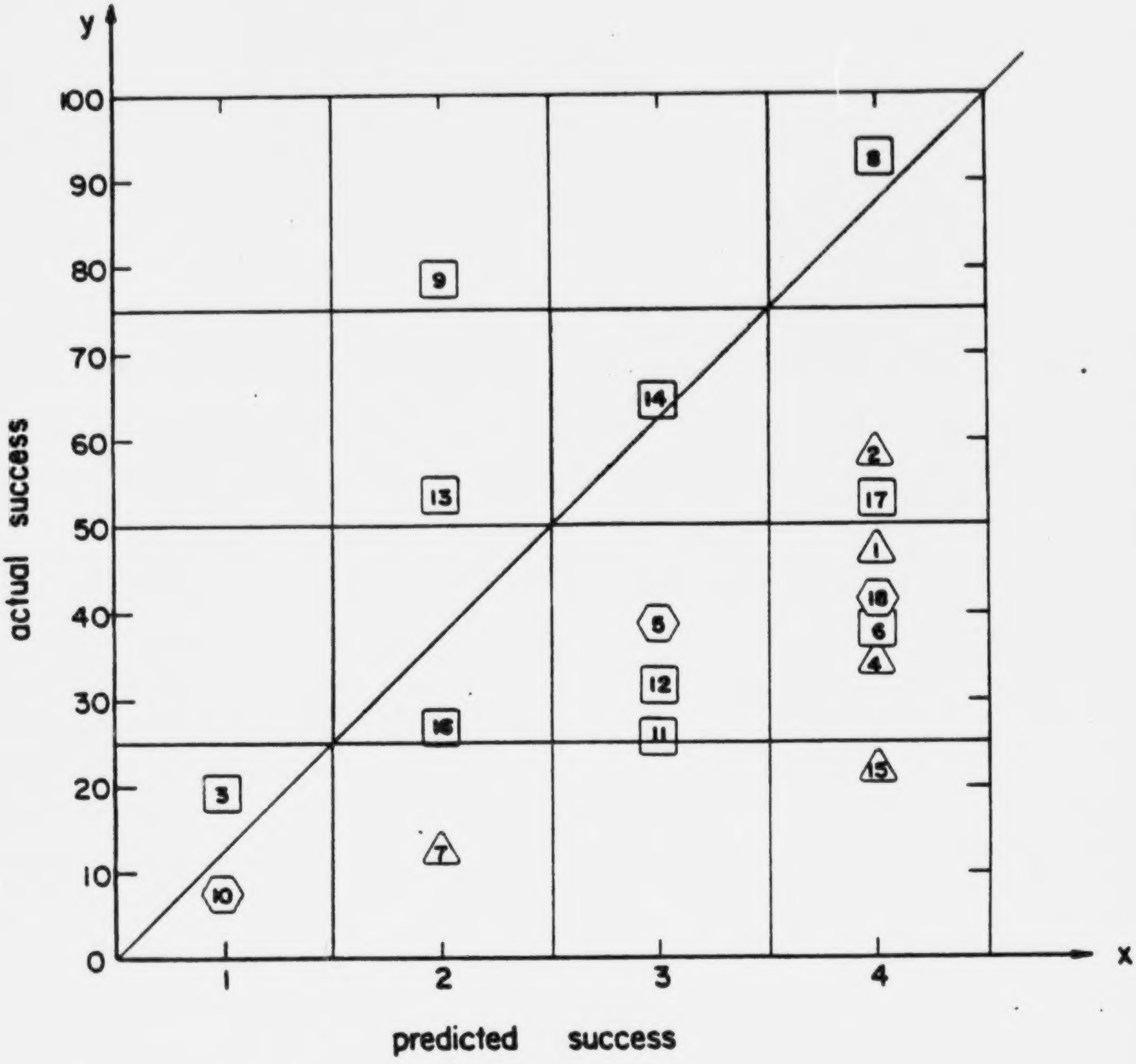


Fig. 6. Graphical representation of data for class (c)

Most questions are classified as comprehension, but the teacher consistently overestimates success (mean score 47). In fact, class observation indicated that the teacher lacked confidence in his own mathematical ability. Although he tried hard he had difficulty in planning and organising a lesson satisfactorily. In this sort of situation, effective teacher reaction to student responses is unlikely.

The same test (with only minor changes) was given to the same classes a second time after about four months. Almost identical pictures were obtained, which would seem to give some validity to the method.

Conclusion

We have described a method of evaluation which permits some penetration into the teaching/learning process, and which connects the three constituents - student, teacher and program. The starting point of the method is subjective estimation by the teachers. Then the test is applied to the class and the information is organised in graphs. From the graphs hypotheses and possible alternative explanations of the situations depicted are proposed. Subsequently, we return to the field and analyse what is happening in reality. It is worth remarking that in the application of the method described in this paper, one of the alternative explanations proposed fitted the reality in the majority of cases. In our curriculum implementation project, the graphical picture is also fed back to the teacher-tutor, who visits the school and finds this useful in guiding and concentrating the direction of his advice to the teacher.

The conclusions reached by this method can be applied to improving teaching strategies, in-service training and the redevelopment of the curriculum materials. The method has considerable potential and can also be applied to many practical aspects of curriculum development and evaluation. For example:

1. Using the analysis of single questions over all classes in a sample, it is possible to design a test which will provide information on problems connected with the teaching of a particular topic.
2. Using the analysis of one class over all the questions, it is possible to try to find characteristics of teachers and teaching strategies which are related to student achievement.

REFERENCES

Avital, S.M. & Shettleworth, S.J. *Objectives for Mathematical Learning.* (Bull. No. 3) Toronto: Ontario Institute for Studies in Education, 1968.

Bloom, B.S. (Ed.) *Taxonomy of Educational Objectives.* Handbook 1: Cognitive Domain. New York: D. McKay Co., 1956.

Bloom, B.S., Hastings, J.T. & Madaus, G.F. *Handbook of Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill, 1971.

Appeared in :

Preservice and Inservice Training of Science Teachers,
Edited by P. Tamir, A. Hofstein and M. Ben-Peretz.
Balaban International Science Services, Philadelphia,
Rehovot, 1983, 573-579.

TEACHERS' EXPECTATIONS VERSUS STUDENTS' PERFORMANCE - A GUIDE TO SOME IN-SERVICE ACTIVITIES

N. Zehavi and M. Bruckheimer

Department of Science Teaching
The Weizmann Institute of Science
Rehovot, Israel

Teachers' expectations versus students' performance - a guide to some in-service activities

A relatively new approach (Clark and Yinger, 1979) to research on teaching assumes that what teachers do is affected by what they think. One of the main concerns in such research is teacher judgement. From the small number of studies on judgement, it is clear that teachers vary as to the accuracy of their prediction of student achievement. Research toward understanding why teachers behave as they do, may bring research on instruction and the behavior of teachers, together with that on curriculum and material.

In a review by Hook and Rosenshine (1979), of what is known about the accuracy of teacher reports of their classroom behavior, one of the conclusions is that teachers do not have practice in estimating their own behavior and then checking against actual performance. More recently Barr (1982) reported on student improvement in a study, in which teachers compared their expectation of student performance with actual performance, and thus became more aware of student difficulties.

Studies in this direction have been conducted since 1978 in the department of Science Teaching at the Weizmann Institute (Zehavi and Bruckheimer, 1981). The department is responsible for the development and implementation of junior high school mathematics curriculum. Teachers' expectations were compared with their students' performance. In this paper, we shall describe a method for analyzing such data in a way that gives some guidelines for in-service activities.

The Method

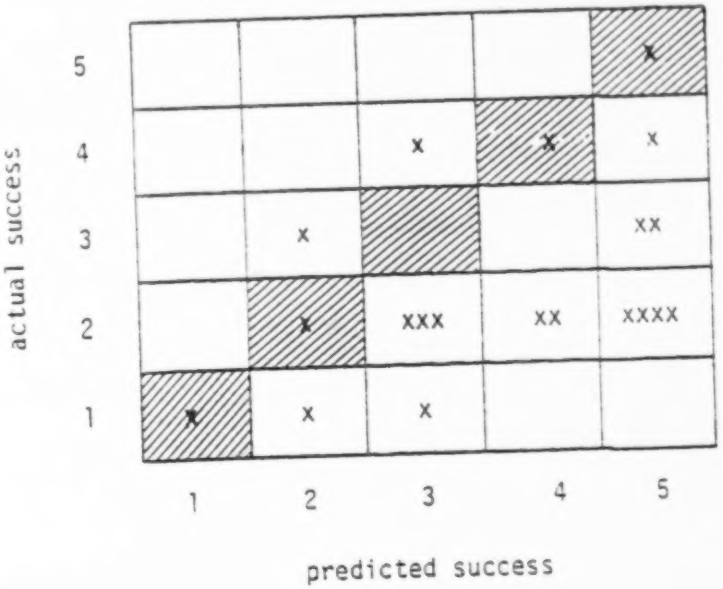
The basic premise of the method is that some of the concepts and parameters used in evaluation are essentially not objective, but subject to the interaction of student, teacher and program. We start by asking the teachers to predict the success rate of their students on each item in a test related to the curriculum. A five point scale is used, from 1 (less than 20% will answer correctly) to 5 (more than 80% will answer correctly).

The teachers are also asked to assign to each item a subjective cognitive level. Following the Bloom taxonomy as modified by Avital (1968) we used the following levels: technique, comprehension and application, analysis and synthesis. After the teachers have exercised their judgement, the test is applied to the classes. The data is then organized graphically to illustrate two types of analysis:

(a) analysis of one class over all the items;

(b) analysis of one item over all the classes.

In general, the teachers in our studies, predicted their students' success on most items with a reasonable degree of accuracy. An exceptional situation is illustrated in Figure 1. The teacher consistently overestimated his class.



	1	2	3	4	5
5					▨ x
4			x	▨ x	x
3		x	▨		xx
2		▨ x	xxx	xx	xxxx
1	▨ x	x	x		

Figure 1: Graphical representation of data for one class over 20 items

Class observation in this case indicated that the teacher had difficulties in planning and organizing a lesson and also lacked confidence in his own mathematical ability. He was gently directed towards attending our summer in-service courses, in which teachers work on the mathematics and the methods relevant to the program.

In another exceptional case, in which the graph depicted severe underestimation, we found a "good" class and a "good" teacher. The class did very well on the test and the teacher was an experienced one. But he was sure that the best way to teach is explain and explain again. He did not believe that his students were capable of doing anything without his first explaining every detail. When he became aware of the situation represented by the graph, he realised that he could and should provide his students with more challenge reflecting their ability.

The information (type (a))obtained from the graphs is fed back to the teacher by his teacher-tutor who visits in the school. They discuss the results, look for explanations in the case of those items which were over- or underestimated and their implication in the teacher's classroom. Teachers who have an "investigative spirit" can use the methods by themselves and may ask for specific guidance. The method itself encourages the teachers to become investigative. An aspect not yet explored, can also serve the researcher: that is, he can use the method to investigate teacher characteristics related to expected actual success data.

Type (b) analysis gives an immediate general information to the curriculum developers. For example, Figure 2 is a graphical representation of the data for one test item over 28 classes. In this case many teachers underestimated their students' success.

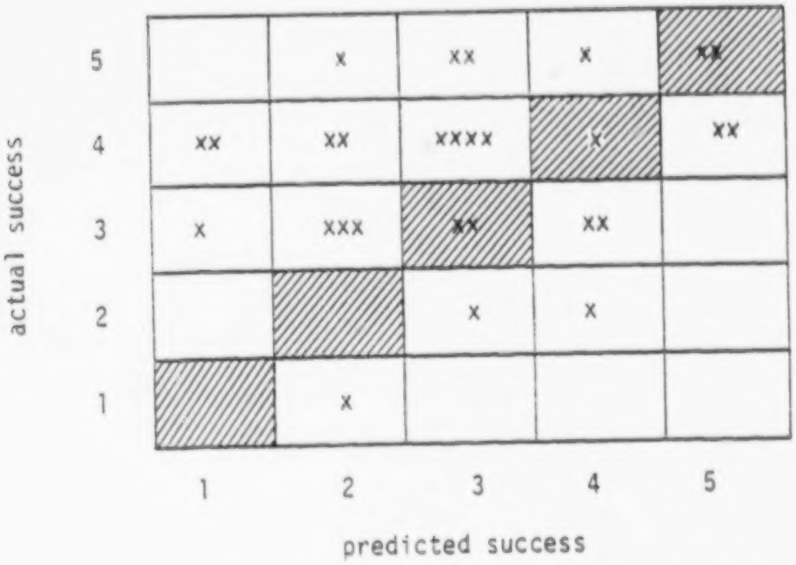


Figure 2: Graphical representation of data for one item over 28 classes.

This item deals with the truth set of an inequality. It was not difficult for the students (as one can seen from the graph). Talking to the teachers we found that they have some "mental block" with regard to this question. We heard arguments like "when I was at school we didn't do sets or inequalities".

Other examples also identified topics in which mismatch was most prevalent. In general, these topics were related to "modern" mathematics and "new" instructional methods. This indicated that in such topics teachers needed still greater help in order to achieve a better understanding of student difficulties.

Further application of the method

A further application of the basic method was designed to verify the suggestion mentioned above; i.e. that there is consistent overestimation or underestimation of items by the teacher newer topics in the curriculum. It is not enough to test such a topic alone, because, maybe teachers tend to overestimate familiar items and underestimate unfamiliar items. In order to be sure that this is not the only cause, one needs to compare two topics to see if there is a difference in the expected-actual success data.

The junior-high algebra curriculum includes both "modern" and "traditional" topics. Thus, the first chapter in ninth grade algebra is "modern" and deals with general function concepts. We decided to apply the method, in order to obtain a comparison between this chapter and one from the same course that can be loosely described as "traditional". One subtest on each topic was prepared. The items in the two subtests were chosen from a bank of items which, in several previous applications had been found to be stable with respect to difficulty and discrimination. The two subtests were balanced for difficulty and familiarity. (Our experience with teacher judgement suggests, that when teachers assign a cognitive level to an item, they are mainly concerned with degree of student familiarity. Thus we asked them to classify items on a three-point familiarity scale: 1-familiar, 2-partially familiar, 3-unfamiliar). Each subtest contained six questions, chosen such that each square in the following matrix contained one question:

difficulty familiarity	easy	hard
familiar		
partially familiar		
unfamiliar		

Figure 3: Subtest design

In 1980, the test was given to 31 classes, for three of which complete data was not obtained. In five classes there was severe overestimation and in a further two exceptional overestimation. As remarked in previous applications of the method, these exceptional cases present interesting case studies, but when we are concerned with the analysis of items over the classes they may introduce bias. Thus these 7 exceptional classes were excluded from the further analysis.

Results

For each item the mean achievement $\overline{y}$ and mean familiarity $\overline{z}$ over the 21 classes was calculated. The values in Figure 4 show that the outcomes fit the design.

	Topic 1 "new"		Topic 2 "traditional"	
difficulty familiarity	easy	hard	easy	hard
familiar	$\bar{y} = 70$; $\bar{z} = 1.1$	$\bar{y} = 49$; $\bar{z} = 1.2$	$\bar{y} = 72$; $\bar{z} = 1.3$	$\bar{y} = 51$; $\bar{z} = 1.5$
partially familiar	$\bar{y} = 63$; $\bar{z} = 2.0$	$\bar{y} = 39$; $\bar{z} = 2.2$	$\bar{y} = 60$; $\bar{z} = 2.0$	$\bar{y} = 37$; $\bar{z} = 2.1$
unfamiliar	$\bar{y} = 64$; $\bar{z} = 2.8$	$\bar{y} = 42$; $\bar{z} = 2.8$	$\bar{y} = 55$; $\bar{z} = 2.7$	$\bar{y} = 32$; $\bar{z} = 2.7$

Figure 4: Mean achievement ($\bar{y}$) and familiarity ($\bar{z}$) for test items in the two topics.

A comparison of the profiles of the two topics based on expected-actual data is shown in Figure 5.

	Topic 1 "new"		Topic 2 "traditional"	
difficulty / familiarity	easy	hard	easy	hard
familiar	over-estimation	over-estimation	precise estimation	precise estimation
partially familiar	under-estimation	over-estimation	precise estimation	under or overestimation
unfamiliar	under-estimation	under-estimation	under or over-estimation	under or over-estimation

Figure 5: Comparison of the profiles of the two topics based on expected-actual data.

The situation for topic 2 is as may be expected. The three easy familiar items are estimated precisely by the teachers and the other three are over-estimated or underestimated almost equally. It is not unreasonable that, for unfamiliar questions on a familiar topic, some teachers will overestimate the outcome and others will underestimate, reflecting individual teaching methods and personality.

The results for topic 1, on the other hand, display a mismatch: the familiar is seriously overestimated and the reverse for the unfamiliar. Although the topic has featured in the textbook for more than ten years and (hence) students still feel uncomfortable and are "out of touch" with each other and with the program. It also may be that the right balance of content and exercises may not yet have been found. For example, some of the relatively sterile exercises of the "domain, codomain, into-onto, etc." variety may profitably be replaced by more demanding and mathematically useful material on the composition of functions, image sets of functions, etc.

There is no doubt that this particular application of the method for analysing groups of test items, indicates that there may still be something to be done in the development and implementation of the newer topics in the curriculum - and that, in spite of an extensive and long-starting in-service training and guidance programme which only again indicates that the time and effort needed to properly implement curriculum changes is far in excess of anything that most curriculum developers have ever realised.

Conclusion

In this paper we have described three types of analysis based on the same principle; that is, the comparison of teacher expectation with actual student achievement. Given a questionnaire devoted to some part of the curriculum we can compare predicted success with the actual outcomes.

If we are interested in the interaction between a single teacher and his class, then the graph depicting expectation versus reality makes immediate visual impact and can focus the discussion. The teacher can also use the method on his own as a form of self-analysis of his teaching. In either case we have the implementation of "research" findings in the teacher's own classroom.

If we concentrate on one item over a number of classes and find that there is a fairly consistent mismatch, this can become the start of a more extensive investigation for possible causes. One possible follow-up, is to use the same method to examine a whole topic (or question type) from which the mismatch question was taken. We exemplified this approach in the comparison between a "new" and "traditional" topic in the junior maths curriculum. The resulting conclusions are also of immediate relevance to the teacher, but he is unlikely, in the majority of cases, to be able to take necessary remedial action unaided. A more likely consequence of such an investigation, is its use in the determination of curriculum renewal, implementation and inservice training strategies.

We have not as yet done controlled studies of what can be done about such cases of mismatch, but the indications are that the analysis can by itself be helpful, as well as identifying priority areas for other inservice activities.

References

Avital, S.M., & Shettleworth, S.J. Objectives for mathematical learning. (Bull.No.3) Toronto: Ontario Institute for Studies in Education, 1968.

Barr, G. Learners, teachers and mathematics. Proceedings of the Sixth International Conference for the Psychology & Mathematical Education. Antwerp, 1982, second volume, 97-101.

Clark, C.M. and Yinger, R.J. Teachers' thinking. In Peterson, P. and Walberg, H. (eds.) Research on Teaching: Concepts, Findings and Implications McCutchan, Berkeley, California, 1979.

Hook, C. and Rosenshine, B. Accuracy of teacher reports of their classroom behavior. Review of Educational Research, 1979, 49, 1-12.

Zehavi, N. and Bruckheimer, M. A Method of Analyzing tests using the teacher's predictions. Journal for Research in Mathematics Education, 1981, 12, 142-151.

בדיקת תרומתן של עבודות סיכום
לפעילות המתמטית של תלמידים המסיימים את חטיבת הביניים

תקציר

נורית זהבי

העבודה המתוארת בתזה זו נועדה לחקור ולפתח פעילות מתמטית במטרה להעלות את התלמיד ואת ההוראה לרמות גבוהות יותר של ההיררכיה הלימודית.

השראה ראשונית שאבנו מרעיונותיו של Polya על למידה היוריסטית, מודל ההערכה האיכותית של פעילות מתמטית מבוסס על עבודתו של Skemp בתחום הפסיכולוגיה החינוכית.

בתחילה פותחו *עבודות סיכום* במתמטיקה לתלמידים בכיתה ט' בתוך מסגרת של מטרות קוגניטיביות וקוגנו-אפקטיביות. עבודת סיכום היא בעיה פתוחה ומקיפה ביותר במתמטיקה,הקשורה ישירות לתכנית הלימודים, מתבססת עליה ומעשירה אותה.

בשל האופי המורכב של החומר והמטרות, אי אפשר היה לעבור מיד לשלב של הערכה משווה. היה צורך לבדוק, האם החומר אכן מסוגל לתרום וכמובן לפתח כלי הערכה רלוונטים למחקר כזה. תוך כדי בדיקת עבודות סיכום של תלמידים נוצרה מסגרת להערכה איכותית. תשובת תלמיד מקבלת ציונים על סקלה של *איכות הפתרון* ועל סקלה של *איכות האמצעים* בהם נבנה ונבדק הפתרון. איכות האמצעים משפיעה בדרך כלל על איכות הפתרון, ואולם תלמיד יכול להגיע לפתרון מלא וכללי ללא הוכחה ומאידך שיקולים ברמה גבוהה לא בהכרח מביאים למציאת התשובה המלאה. את שני הממדים האלה, *פתרון ואמצעים* ניתן לתאר בטבלת *איכות התשובה*. שיפור בפעילות מתמטית יתבטא ב"תנועת האוכלוסיה" בכיוון הרצוי בטבלה כזו. מובהקות סטטיסטית של שיפור אפשר לבדוק על-ידי Wilcoxon-Raatz-test שהוא גירסה משופרת של המבחן הלא-פרמטרי של Wilcoxon. *מבחני משימות* בתוך מסגרת ההערכה שתוארה הוכנו ותוקפו ונעשה בהם שימוש במחקר השוואה בעל מערך של קבוצת ניסוי - קבוצת ביקורת - מבחן pre - מבחן post. מימצאי המחקר מצביעים על תרומה משמעותית של עבודות הסיכום לפעילות המתמטית.

במהלך המחקר התברר כי לצורך הפעלה יעילה של החומר על-ידי המורים נחוץ להכין גם מסגרת מובנית של הדרכה. לשם כך נערך מחקר קליני בכיתה אחת וכל התלמידים רואיינו לאחר שעבדו על משימות. המטרה היתה לגבש קשרים בין קשיי התלמידים ואופי ההדרכה המתבקשת. התוצאה היתה מודל להערכה והדרכה דיאגנוסטיים. המודל שפותח נוסה בהצלחה על-ידי מורים במספר כיתות.

בדיקת תרומתן של עבודות סיכום
לפעילות המתמטית של תלמידים המסיימים את חטיבת הביניים

חיבור לשם קבלת תואר

דוקטור לפילוסופיה

מאת:

נורית זהבי

מוגש אל

המועצה המדעית

מכון ויצמן למדע

רחובות

חשון, תשמ"ה

237

BEST COPY AVAILABLE

בדיקת התרומה של עבודות סיכום
לפעילות המתמטית של תלמידים המצטיינים של חטיבת הביניים

חיבור לשם קבלת תואר
דוקטור לפילוסופיה

מאת:

נורית זהבי

מוגש אל
המועצה המדעית
מכון ויצמן למדע
רחובות

חשון, תשמ"ח